

Courage, vision and the Department of Energy

Federico Peña is heading for one of the hottest seats in the United States government. He will need farsighted resolve to pursue his department's controversial goals.

The secretary of energy is the only one of Washington's elite band of science administrators to enjoy a full place in the President's cabinet, and also the only one to be replaced so far by President Bill Clinton as he embarks upon his second four-year term. But the job seldom allows much opportunity for its occupant to bask in the reflected glory of the department's scientific achievements, which are considerable. Instead, the management of the Department of Energy \$6-billion research portfolio and its other activities tends to involve every energy secretary in protracted political infighting by virtue of the wide portfolio of controversial issues on the department's turf.

The tenure of Federico Peña, who is likely to succeed Hazel O'Leary in the position after his successful confirmation hearings before the Senate last week (see page 472), will doubtless reverberate to a long series of public clashes with the many interests whose representatives feel entitled to tell the energy department what to do. Add to that the recent danger that the Congress would try to abolish the department altogether, and it is no wonder that anyone arriving in this position will do so with low expectations.

Yet Peña enters this particular hornets' nest with some advantages. One is that the abolitionists are in retreat: the Republicans, who still control the Congress, appear to have lost their revolutionary appetite for dismemberment. Another advantage is that the main components of the department are in the hands of seasoned assistant secretaries well able to defend their programmes in the Congress without much help from the secretary's office. That said, if Peña is confirmed, he can be sure that the thorniest issues will land squarely in his lap.

These include reform of the large network of laboratories, including the three associated with nuclear weapons, that the department operates. Two years ago this week, a panel chaired by the industrialist Robert Galvin called for dramatic action to improve their efficiency. The laboratories and their friends in Congress have successfully opposed wholesale reform, but its rapid implementa-

tion would ultimately be in their own best interests.

The stockpile stewardship programme, which will replace nuclear testing with a \$4-billion-a-year research and development programme based on new supercomputers and experimental facilities, has strong bipartisan support and seems unlikely to come unstuck on Peña's watch. An even larger programme to clean up the defunct nuclear weapons production complex will continue to run into various local difficulties at sites across the country. But some necessary cuts in the programme have been made, and improvements in its management should soon start to bear fruit.

The problem that could occupy the largest slice of Peña's attention throughout his tenure, however, is the disposal of spent fuel from nuclear power stations. The electricity utilities have paid billions in levies towards the costs of a permanent store. Next year, the department will complete its assessment of Yucca Mountain, Nevada, as a suitable site. Peña's task is to convince people that he will stand up to both the utilities and the environmentalists and handle Yucca on the basis of what is right for the United States.

O'Leary once said that the energy department should really be called the department of science and technology. But, in fact, the department's central mission today is to manage the tangled legacy of outdated priorities and nuclear waste. O'Leary was sometimes chastised for her lack of attention to detail, but her own track record on this central mission has been impressive. She tried to open up the department to a deeply suspicious public, and she brought round recalcitrant laboratory directors to a world without nuclear tests.

Federico Peña comes to the department with no experience of the relevant technical issues, but with a reputation as a hard worker and something of a visionary. He will be judged not on his grasp of unfamiliar detail but on the courage he shows in confronting special interests, and the vision he can formulate for a department that has yet to come fully to terms with the end of the Cold War. □

Avoid financial 'correctness'

Insistence that authors declare business interests in papers is beside the point.

It comes as no surprise to find (*Nature* 385, 376; 30 January 1997) that about one-third of a group of life scientists working in the biotechnologically rich state of Massachusetts had financial interests in work that they published in academic journals in 1992. Nor, given the absence of policies of most journals at the time, is it surprising that those interests were seldom declared in their papers. More recently, some journals have insisted on declarations of interest, either to editors only or in published papers.

This journal has never required that authors declare such affiliations, because the reasons proposed by others are less than compelling. It would be reasonable to assume, nowadays, that virtually every good paper with a conceivable biotechnological relevance emerging from the west and east coasts of the United States, as well as many European

laboratories, has at least one author with a financial interest — but what of it? The measurements and conclusions are in principle unaffected, as is the requirement that uncertainties be made clear. Would the impact on readers of a declaration of interest differ from that of a statement that the success of the next grant application will be that much greater as a result of publication? And who is to police disobedience? Such appeals for openness are selective — other pieces of information would be just as (ir)relevant to a paper's content.

The work published (*Science and Engineering Ethics* 2, 395; 1996) makes no claim that the undeclared interests led to any fraud, deception or bias in presentation, and until there is evidence that there are serious risks of such malpractice, this journal will persist in its stubborn belief that research as we publish it is indeed research, not business. □



Tunnelling techniques hold key to prospects for giant accelerator

[WASHINGTON] Advances in tunnelling and materials technology could revitalize plans dreamt up in the 1980s to build a vast accelerator ring up to 1,000 km in circumference in the United States, according to physicists at Fermilab — the Fermi National Accelerator Laboratory — outside Chicago.

A small team working on future accelerator concepts believes that new techniques for automated tunnelling, together with the future availability of superconducting cable to power simple iron electromagnets, could make possible the construction of a giant ring able to achieve more than ten times the energy of Europe's planned Large Hadron Collider (LHC).

But, since the Fermilab team began to study the concept seriously in November 1995, it has encountered serious obstacles. Robotic boring devices being developed by civil engineers to plant pipes beneath rivers and streets are unlikely to enhance the project's viability as much as had been hoped.

Another hadron collider is one of the options being explored by physicists as they look beyond the capability of the LHC, which is due to be completed in 2005 at CERN, the European laboratory for particle physics in Geneva, Switzerland.

Rival teams are also developing less exotic concepts for linear electron colliders, and for machines to examine collisions of muons, unstable particles of intermediate mass.

Anyone wanting to design an even larger hadron collider has two options: to pursue a far more powerful magnetic field (as have the designers of the LHC, which is an upgrade to CERN's existing ring), or to look to a much larger ring. Advocates of a larger ring have

named it the Pipetron, and are seeking to develop their concept into a project that could one day gain public support.

The Pipetron could, its advocates believe, use a C-section magnet to focus and gently steer the hadron beam, rather than the more complex combination of bipolar and quadrupole magnets used in more compact accelerator rings. The first prototype of a short section of such a magnet is now being built at Fermilab, which will spend a modest \$200,000 on the concept this year.

"We have the feeling that this is a low-cost approach, but we have to prove that," says Ernie Malamud, who is working on the Pipetron concept at Fermilab. The concept could be viable, he says, because technologies have changed so much since it was previously contemplated, in the early 1980s.

Malamud cites high-temperature superconductors, as well as digital electronics for control and monitoring, as examples. But it was the potential of 'microtunnelling' — the use of remotely controlled machines to dig small-bore tunnels cheaply — that Pipetron advocates hoped would lead to significant cost breakthroughs.

"The community is looking for ways to lower the cost of expanding the energy frontier," says Malamud. An added advantage of the Pipetron concept, he says, is that the drilling technology developed to build it might help public utilities to put cables and pipes underground, as the public wishes.

Industrial corporations are showing interest in the concept, which they view as an opportunity to develop at least three important technologies. Paul Grant of the Electric Power Research Institute, an electric utilities



Ring road: Pipetron would form a tunnel of up to 1,000 km in circumference beneath the mid-west.

research consortium based in Palo Alto, California, says that the project could help to pioneer automated tunnelling, superconducting cable for power transmission, and the use of robots for construction and maintenance in small tunnels.

Physicists at Fermilab foresee a machine of up to 1,000 km in circumference, dug in a consistent slab of dolomite rock and extending from the laboratory site. Fermilab's director, John Peoples, stresses that the Pipetron is only one of the approaches the laboratory is considering for its own future after the LHC supplants Fermilab's Tevatron as the world's most powerful hadron collider. The main problem, he says, is cost.

Peoples says that costs per TeV would have to come down by an order of magnitude from those of the ill-fated Superconducting Super Collider, abandoned in Texas in 1994, if a machine is to have a chance of being built. He believes that automated boring of small-bore tunnels followed by automated operation "with no human access" is needed to achieve such a cost breakthrough.

But, after consulting tunnelling engineers, the Pipetron team is becoming less sanguine about automated tunnelling or drilling. "We're told it's cheaper to be bigger," says Malamud. "We got talked out of the original idea of a non-accessible tunnel."

According to Jim Friant, a Seattle-based tunnelling consultant who has advised the Pipetron team, today's cheapest tunnels are 14 ft in diameter. This could be brought down to 8 ft, he says. With anything smaller, men cannot get to the front to replace cutters while the machine is in place.

Small, remotely controlled tunnel cutters are liable to get stuck, and are hard to steer. The concept of a robot that could dig the Pipetron cheaply appears, for now, to be merely a pipe dream.

Colin Macilwain

Prime minister heads Indian science panel

[NEW DELHI] India's prime minister, H. D. Dewe Gowda, is to head a committee on science and technology that has been set up to help accelerate research and the application of technology to national development. The other members of the panel will be cabinet ministers in charge of science, finance, agriculture, defence, industry, rural development and education.

The government has also revived the scientific advisory committee to the cabinet which became defunct in 1989 after the death of the then prime minister, Rajiv Gandhi. It will be made up of leading scientists from government and industry, headed by an eminent scientist.

While the committee will formulate

policies and specific programmes, the cabinet panel will provide the political backing needed to translate its proposals into action. A committee made up of the secretaries of the various science-funding departments, headed by the cabinet secretary, will be responsible for implementing the programmes.

Ragunath Mashelkar, secretary to the Department of Scientific and Industrial Research, says that the setting-up of these committees is a signal that funding prospects for Indian science may improve in the next financial year. India's total spending on scientific research fell from 1.1 per cent of its gross domestic product in 1990 to 0.83 per cent in 1996.

K. S. Jayaraman

Entire *E. coli* genome sequenced — at last

[LONDON] A long-running project to sequence the entire genome of the bacterium *Escherichia coli* has finally reached its goal — at first glance, twice over. Two independent groups, one headed by Fred Blattner in the United States and the other by Takashi Horiuchi in Japan, made the announcement last week, seven years after the project began.

Each group has placed its sequence data in public databases. But while Blattner has issued the complete annotated genome of a single *E. coli* strain, K-12, the unannotated Japanese sequence is a hybrid of more than one strain, and up to 35 per cent of it appears to be drawn from Blattner's data.

The road to completing the *E. coli* sequence has been long and bumpy. When the sequencing project was announced in 1989 by the Japanese Ministry of Education, Culture and Science, and by Blattner at the University of Wisconsin at Madison, it was the first to be aimed at an independently living organism.

Thousands of researchers around the world were already sequencing fragments comprising some 10 per cent of the genome because of the bacterium's key role in biological research. But three smaller projects that had begun much later using advanced technology to sequence the genomes of *Haemophilus influenzae*, *Mycobacterium genitalium* and *Methanococcus jannaschii* have already beaten the 4.6-megabase *E. coli* sequence to the finishing line.

Those three smaller projects were carried out at the Institute for Genomic Research in

Unravelling? Two groups of researchers have independently sequenced the *E. coli* genome.

Rockville, Maryland. Craig Venter, director of the institute, says the *E. coli* sequence "is a very significant accomplishment, made much more so by the changes in technology and approach since that project started".

Blattner's work depended on methods that he had invented out of necessity. "From the beginning it was a matter of adapting the technology currently available," he says. "In that sense, Blattner started too early," says Manfred Kröger, of the Institute of Microbiology and Molecular Biology at the University of Giessen, Germany, who since 1989 has maintained a public database collated from all the released sequencing data.

Initially, Blattner was unable to get funding for his preferred method of 'whole genome' sequencing, and funding agencies did not at first recognize the necessity for automated sequencers. But in July 1995

when the US National Institutes of Health offered grants to complete the sequence in open competition, Blattner won the money and rebuilt his research group. He brought in five new high-throughput (33-lane) Applied Biosystems sequencing machines to march through the remaining nucleotides. During 1996 alone, the team sequenced 2.8 megabases by the completion date of 19 December.

Meanwhile, by 1994 it had become clear that Japan's independent project based on a restriction map of the genome, and shared by an 11-institute consortium, was far behind schedule (see *Nature* 368, 383; 1994). Horiuchi, at the National Institute for Basic Biology at Okazaki, was appointed leader in April 1995 and extra money was granted to ensure completion of the sequence "as a matter of national pride", says Kröger. So far, however, only the raw sequence data are available in the public database.

Venter, who had predicted that Blattner's genome sequence might not feature among the first ten to be released, says he is "pleased that he proved me wrong". Blattner comments: "Don't forget that *E. coli* is larger than the [Rockville] genomes combined".

According to Venter, preliminary (unpublished) analysis of the sequence indicates that it contains many genes of unknown function. But Gordon Dougan, professor of biochemistry at Imperial College, London, says that the detailed sequence will be a useful guide to "the evolution of virulence" in pathogenic strains of *E. coli* and of other enteric organisms.

Claire O'Brien

Peña plays it straight to keep on course for US energy secretary

[WASHINGTON] Federico Peña's appointment as head of the US Department of Energy seems virtually certain after he emerged unscathed from a lengthy confirmation hearing in Washington last week.

Before the hearing, Peña, previously a secretary of transportation, had been warned by Senator Frank Murkowski (Democrat, Alaska), the chairman of the energy committee which held the hearing, to expect a tough grilling on issues such as nuclear waste. The energy department has admitted it is likely to violate a court order to begin accepting spent nuclear fuel from US utilities from the beginning of next year.

Avoiding detailed responses that could get him into trouble later, Peña did not stray from the Clinton administration's position of rejecting legislation to establish an interim federal nuclear waste store near Yucca Mountain, the Nevada site that may become a permanent waste repository.

Peña gave no indication of any departure

from the policies of his predecessor, Hazel O'Leary. He said he would oppose proposals being considered in Congress to downgrade the agency's status, given the advantage of being able to sit "at the table with the president", or to dissolve the department, distributing its responsibilities to other agencies, chiefly the defence department.

Peña praised the energy department's system of national laboratories, calling them the "crown jewels of US scientific leadership". He promised to continue responding to the recommendations of the 1995 Galvin Commission — which urged the department to slash regulations and paperwork at the laboratories — and to try to establish stronger partnerships between the laboratories and the private sector.

The toughest questioning came from Jon Kyl (Republican, Arizona), who argued that the agency will violate a law signed by President Bill Clinton last year if it fails to decide by the end of 1997 which technology it

will use for producing new stocks of tritium.

The agency has said it will decide in 1998 between using an existing commercial light-water reactor or building a linear accelerator, in order to begin producing tritium for the nuclear arsenal by 2005. Peña said he was unaware of the requirement to make a decision this year.

Despite strong protests from environmental and disarmament groups, Peña said he will proceed with a series of "subcritical" underground nuclear experiments planned for this year.

He also assured senators that he will seek adequate funds for the agency's ambitious effort to develop experimental surrogates for testing, including a series of increasingly powerful supercomputers and the \$1.1-billion National Ignition Facility for inertial confinement fusion. Peña said that the administration's budget request for 1998 will include \$4 billion for "[nuclear] stockpile stewardship" activities.

David Kramer

UK calls a halt to moves to privatize laboratories

[LONDON] A wave of relief tinged with irritation flowed through many British government laboratories and research organizations last week at the news that the government has, after a lengthy review, decided not to transfer them to private ownership.

Among the organizations that will remain in the public sector are the British Geological Survey, part of the Natural Environment Research Council, and the National Radiological Protection Board. The news also affects several units run by the Medical Research Council (MRC), although one — the Virology Unit — is likely to be taken over by the University of Glasgow.

The decision on 37 separate laboratories represents the outcome of the most recent stage of the so-called 'prior options' review. This was launched by the Conservative government in 1993 to review all government-run laboratories and examine whether they should continue to be funded directly by the public sector.

Many scientists whose futures had been hanging in the air for many months and who had feared widespread redundancies if the privatizations had gone ahead were expressing relief at last week's news.

There was also relief within the research councils concerned. These will not now, as had been threatened, have to find large lump sums representing the 'crystallized' pension rights of staff being transferred to private employers out of their research funds. (The Treasury had been refusing to make extra money available.)

They will also be able to go ahead with plans that had been held up while the review was under way. These include the appointment of new directors to the MRC's Reproductive Laboratory Unit in Edinburgh, and the Dunn Nutrition Unit, which will be 'reconstituted' on moving into a new building next to the university clinical school in Cambridge.

But there was also irritation that results of the prior options exercise, carried out at considerable expense and requiring a substantial effort from the laboratories involved in preparing reports and opening themselves up to assessment, appears in many cases to have resulted in little more than a call for more efficient management.

"This whole process has been dragging on for a long time, and I am pleased to see that the government has realized the problems that it was causing in terms of the loss of morale, and has decided to finish off the matter," says Aaron Klug, president of the Royal Society and the former head of the

MRC's Laboratory of Molecular Biology in Cambridge. "Questioning things is always useful, but not if it involves costly and lengthy procedures."

Klug says that the society, which has expressed concern at the dangers of privatizing bodies that form much of Britain's scientific infrastructure, welcomes the government's announcement that it will in future carry out regular reviews of individual laboratories. "We have been seeking that for a long time. All institutes need to be reviewed to prevent the accumulation of dead wood."

The move to privatize broad chunks of this infrastructure had been promoted by senior government figures, in particular Michael Heseltine, the deputy prime minister. As the former President of the Board of Trade, it was Heseltine who pushed through the transfer to private management of bodies such as the National Physical Laboratory.

But others, including, reportedly, some officials at the Office of Science and Technology — which now forms part of the Department of Trade and Industry — are said to have become increasingly convinced that, whatever the ideological principles involved, the advantages of privatization were outweighed by the potential costs and disruption.

Valerie Ellis, assistant general secretary of the Institution of Professional Managers and Specialists — the labour union that represents many government scientists — says that the decision "vindicates our belief that the whole exercise has been a waste of time and money".

She says that an important factor that may have helped to persuade the government against widespread privatization of its scientific services was its experience over the outbreaks of bovine spongiform encephalopathy and *Escherichia coli* food poisoning, both of which underlined the need for access to top-level scientific advice.

Labour has promised that if, as is widely expected, it comes to power in the forthcoming general election, it will halt the prior options process. Adam Ingram, shadow minister for science and technology, said he welcomed the government's decision to "back off" under pressure from scientists.

But given science's funding difficulties, a Labour government is still expected to have to find ways in which the laboratories can be run more effectively.

If the Conservatives make a surprise comeback, many expect that the privatization campaign will start again. But, with an election looming and pressing decisions being held up, "basically they just ran out of time," says Ellis.

David Dickson

First national review in Australia looks at 'gaps and overlaps'

[SYDNEY] A wide-ranging review that could eventually lead to a major reorganization of government-funded science and policy advice has been quietly launched in Australia. Although individual agencies have experienced repeated reviews and restructuring over the past decade, this is the first exercise to look at the national situation.

The review is being carried out by John Stocker, who was recently appointed part-time chief scientist (see *Nature* 384, 300; 1996). Peter McGauran, science and technology minister, has asked Stocker to report on "any gaps and overlaps in current arrangements" as well as "ways of identifying national priorities".

Stocker says the idea for the initiative came from the National Commission of Audit, set up by John Howard, the Conservative Coalition Prime Minister, after he took office last March. The commission concluded that, with at least seven government portfolios having an interest in science, "there



Stocker: reporting by mid-1997.

appears to be considerable scope for rationalisation, improved efficiency and the achievement of more targeted outcomes".

Australia has lacked a science department since it was abolished by the Labor government more than a decade ago.

McGauran is responsible for the activities of the Commonwealth Scientific and Industrial Research Organization (CSIRO), the Australian Institute of Marine Science and the Australian Nuclear Science and Technology Organization, but not for any of the grant-giving organizations.

Nevertheless, the statutory bodies covering marine and nuclear science, as well as the Australian Geological Survey Organization, are prime targets for being taken into CSIRO. But they will probably not go willingly.

Some advisory bodies could be merged. But the main grant-giving bodies — the Australian Research Council, National Health and Medical Research Council and the Industrial R&D Board — have solid support within the education, health and industry departments respectively, and are likely to argue strongly against being centralized.

Unusually for national inquiries, the review has started without any public announcement, requests for submissions or public hearings, and is being conducted by one person. It will be completed quickly. Stocker has already contacted key individuals, and says that he has to report to McGauran "by mid-year".

Peter Pockley

UK seeks to appease food safety critics

[PARIS] The British government last week announced plans to set up an independent food safety council to advise it on food policy, after conceding that it is no longer trusted by the public on such issues in the wake of the 'mad cow' crisis.

The proposal has been welcomed by farmers' groups as a way of addressing public concern about safety. But consumer groups and opposition parties have criticized the government for stopping short of setting up a food agency independent from the Ministry of Agriculture, Fisheries and Food.

The council, whose creation will depend on the Conservatives winning the impending general election, would have 20 expert and lay members and would meet four times a year. It would be free to comment on all aspects of government food policy, such as safety, quality and labelling. The conclusions of the council's meetings would be made public, and it would submit an annual report to parliament.

The government has also announced that it plans to appoint a Food Safety Adviser, a senior official similar in status to its Chief Medical Officer, to head the council.

The part-time adviser would be an eminent scientist who would answer directly to ministers, but who would in principle be free to criticize government policy. He or she would have an important role in dealing with

the media and appearing before parliamentary select committees to explain the council's position on government policy.

The plans were unveiled last week by the agriculture minister, Douglas Hogg, and the health secretary, Stephen Dorrell. Hogg said the council would leave executive control of food policy to government but would act as a watchdog free to "speak out robustly" on government policy. "We are making a rod for our own backs," he said.

Hogg conceded that the public no longer believes what it is told by the government on food safety issues, perceiving it to be "too closely related with the producing industry". He said he hoped that the council would "greatly enhance people's perceptions of the independence of the advice that government is receiving".

The proposals have been cautiously welcomed by farmers' groups. "This model could be made to work if it has the right people with autonomy," says Ian Gardiner, director of policy at the National Farmers' Union.

But Sue Todd, an official at the UK Consumers' Association, says that, while the association welcomes the government's recognition "that there is a problem", the proposals fall far short of what is needed to restore consumer confidence in government decision-making after the crisis over bovine spongi-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Hogg: Food safety council would be free to "speak out robustly" on government policy.
form encephalopathy (BSE).

The council would be "just another advisory body", claims Todd. She argues that its creation fails to address the fundamental problem: the conflict arising from the fact that the agriculture ministry is responsible for both promoting the interests of industry and protecting consumers. She says that in practice consumer interests tend to lose out.

Todd argues that the government department that promotes agriculture and food production needs to be separate from the regulatory body, as is being proposed at the European Commission (see *Nature* 385, 285; 1996). This separation would encourage a more open debate "about how decisions are made and the evidence on which they are based", she says.

Similarly, the opposition Labour party, which has promised to set up an independent food agency with regulatory powers if it wins the election, says the council would lack "the powers or resources to make a major impact".

The fact that the government has stated explicitly that a major role of the new adviser would be to handle the media and general public has also raised concern that the post may have as much to do with public relations as public safety.

Meanwhile, a report by the French senate last week called for the setting up of a national food safety agency separate from the French ministry of agriculture, as well as the creation of an early-warning body similar to the US Centers for Disease Control in Atlanta.

The author of the French report, Claude Huriet (Meurthe-et-Moselle), complained of glaring gaps in France's institutional capacity to manage public-health risks, and deplored the low funding for such activities compared with that in the United States.

An independent European Union food agency or even a world food agency may be needed, says Gardiner. He argues that the gradual liberalization of trade in food and agriculture is likely to see an increase in disputes such as those about genetically modified organisms, or the use of hormones to boost beef and milk production. **Declan Butler**

AIDS 'cure' scientists go to top for funds

[CAPE TOWN] Three scientists at the University of Pretoria, who claim to have found a cure for AIDS, last month appealed to the South African cabinet for R3.7 million (US\$800,000) to continue their research.

But their use of such an unconventional channel has focused attention on both the product itself, patented as Virodene P058, and its clinical trials, which are alleged to have been carried out without going through normal approval procedures.

The interview with cabinet members was arranged by the health minister, Nkosazana Zuma, who appeared reluctant to take personal responsibility for granting the scientists' request for funding from her department. The department has R40 million available, from the European Union and South Africa's own Reconstruction and Development Programme funds, for research and other activities aimed at combating sexually transmitted diseases.

The cabinet gave a standing ovation to the three researchers — Dirk du Plessis and Kallie Landauer, both cardiothoracic surgeons, and Olga Visser, a technician. The deputy president, Thabo Mbeki, said their request would be considered. But the following day the University of Pretoria announced that a

committee had been set up to investigate a "deviation from the university's established research practices" by the three researchers.

It is alleged that phase one clinical trials were carried out on 12 patients without the consent of either the university's ethics committee or the Medicines Control Council, the authority that considers applications to conduct drugs trials.

Peter Folb, chairman of the council and professor of pharmacology at the University of Cape Town, confirms that it had not previously reviewed an application for testing Virodene P058, but that an application had been rejected by the University of Pretoria's ethics committee.

The council has now started its own investigation into the pharmacological properties of the drug. This is due to be completed this week, as is the university's inquiry. The drug is said to contain a compound that can penetrate lymphocytes and inhibit viral replication because of its low molecular weight and its composition. There has been speculation that Zuma herself granted permission for clinical trials, bypassing both the university and the national authorities. But she has been unavailable for comment. **Michael Cherry**

University teaching hospitals battle for scarce funds in Berlin

[MUNICH] The leading hospital in the former East Germany, the Charité in Berlin, is likely to receive priority over the next few years in the redevelopment of the city's teaching hospitals. In a statement last week, the Wissenschaftsrat, Germany's science council, recommended focusing financial support on the Charité, which is the medical faculty of Humboldt University in east Berlin.

But that could result in the Free University's Benjamin Franklin Hospital in west Berlin ending up on the scientific scrapheap, as the city — which is almost bankrupt — seems incapable of paying for the urgently needed renovation of the faculty.

After German reunification, Berlin was faced with a surplus of university departments and teaching hospitals. Since then, the Charité has completely transformed itself. After an evaluation, only 20 per cent of its professors were

reappointed, most of the remaining posts being filled with scientists from the west. In 1995, after bitter disputes (see *Nature* 369, 431; 1994), the Charité merged with the nearby Virchow Hospital, previously part of the Free University.

Since then, seven departments have been closed and the number of beds cut by 500 to 2,200, in line with a law on the restructuring of university medicine passed by the city government in 1994. "Our radical policy has been a success," says Wolfram Sterry, head of Charité's medical faculty. "As regards research grants and scientific publications, we are already among the top third of German teaching hospitals."

The Wissenschaftsrat has now rewarded these achievements. While DM400 million (US\$245 million) has already been used for building work at the hospital since 1994, the science council's support means that a further DM400 million is likely to be granted by the city and federal governments

to renovate the Charité's dilapidated infrastructure.

But that leaves the Benjamin Franklin Hospital out in the cold. "I cannot understand why the Charité should still be given priority, when we have already handed over the well-equipped Virchow Hospital to them," says Peter Gaethgens, head of the Free University's medical faculty. "In terms of public budgets, this recommendation means that in practice we will get nothing, while Charité takes it all."

Gaethgens says that the two teaching hospitals should be treated even-handedly. Peter Radunski, Berlin's science minister, has promised to support the Benjamin Franklin Hospital. But Gaethgens is pessimistic, saying: "The health ministry seems to be playing the dominant role. They want us to trim back to help cover [the ministry's] financial problems caused by the bankruptcy of Berlin's public health insurance."

Quirin Schiermeier

Mellow mood fosters marijuana research

[SAN FRANCISCO] A Californian senator plans to introduce legislation to provide \$6 million over three years to set up a centre to carry out research into the potential medical uses of marijuana. The research would be overseen by the University of California.

The move coincides with signs that officials in Washington have begun to moderate their hostility towards physicians who respond to a decision by voters in California last autumn to legalize medical uses of the drug (see *Nature* 384, 95; 1996). Federal officials are also showing increasing interest in research in the field.

Proposals for a state-funded research centre were unveiled last week by state senator John Vasconcellos (Democrat, San Jose), chairman of the criminal procedure committee in the state Senate. He said that the centre would conduct clinical trials on the effectiveness of marijuana, study adverse effects and safety, and develop intervention techniques for marijuana abuse.

Rand Martin, chief of staff to Vasconcellos, says the bill was written in answer to the federal government's "shameful" response to last year's vote. He says the federal response has been "more driven by hyperbole and politics than compassion and research needs".

The California law requires only an oral recommendation from a physician, not a prescription. But federal officials have warned doctors in both California and Arizona, which passed a similar law, that they risk federal prosecution if they recommend

the drug. Janet Reno, the Attorney General, has also said that doctors could lose the right to write prescriptions and collect reimbursements for Medicaid and Medicare patients.

Some Californian physicians say that individuals they suspect to be federal agents have been coming to their offices asking questions about marijuana, and have expressed concern at the possibility that threats of government action could lead to the monitoring and censoring of their conversations with patients.

Last month, a group of San Francisco physicians and their patients filed a suit against four federal officials claiming that the federal threats violate their constitutional right to free speech. And last week, an editorial in the *New England Journal of Medicine* said federal policy was "misguided, heavy-handed and inhumane".

Some see a possible softening of the

administration's stance in a decision by the National Institutes of Health (NIH) to hold a scientific workshop later this month to survey the literature on the medical use of marijuana, and to consider the potential for research into the drug's possible effectiveness, as well as its risks.

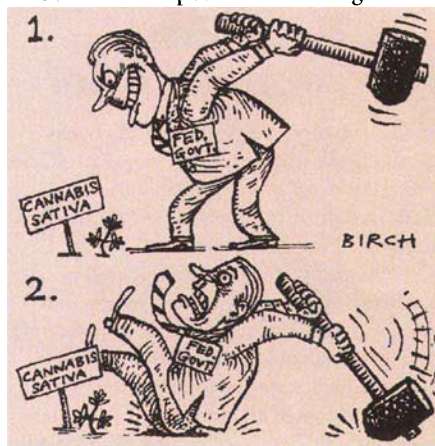
Harold Varmus, director of the NIH, says that the Clinton administration is very responsive to the idea of getting more information on which policy decisions can be made. Explaining the thinking behind the conference, he says that the state referenda have created a responsibility for public health officials to investigate whether marijuana is efficacious, and if so, to develop standard treatment schedules.

Separately, Barry McCaffrey, director of the Office of National Drug Control Policy, said the government would provide \$1 million to the Institute of Medicine to review existing research on the medical use of smoked marijuana.

But the San Francisco Medical Society calls such a study "unnecessarily duplicative and untimely", as the US Drug Enforcement Administration and other bodies have already compiled such data.

Until now, a lack of legally available marijuana for research purposes has blocked even studies approved by the Food and Drug Administration. The California bill would not create such a supply. But Vasconcellos hopes that the state administration will help resolve the problem. Varmus says that the NIH is prepared to provide marijuana for an approved protocol.

Sally Lehrman



Medical network pioneers live 3-D surgical images

[TOKYO] The world's first live test transmissions of three-dimensional medical images of patients have been launched by Nagoya University in Japan and Duke University Medical Center in Durham, North Carolina, in the United States, using a high-capacity trans-Pacific network.

The experiment is part of the Global Interoperability for Broadband Networks project set up by the G7 group of seven leading nations. Several 'telemedicine' projects already exist in Japan linking hospitals and medical research organizations. The Nagoya-Duke network uses a three-dimensional television system developed at Nagoya that allows reconstruction of stereoscopic images of structures in the human body.

According to Tomohiko Hattori of Nagoya University, who developed the system, several doctors can view the three-dimensional images at the same time. Doctors in Japan and the United States can use the network to exchange information about diagnostic and surgical techniques.

The international system is still in its experimental phase. But a telemedicine network developed at the National Cancer Centre in Tokyo is at a more advanced stage.

The national network began operating in April 1994 with a 6-megabit-per-second optical-fibre link between the Tokyo institute and a new campus of the cancer centre in Chiba, 30 km east of Tokyo. It allows transmission of still high-definition television images such as X-ray images and teleconferencing using the standard NTSC format used in US and Japanese televisions, as well as remote operation of a microscope at the cancer institute.

Six other Japanese cancer centres have since joined the network: the National Sapporo Hospital in the northern island of Hokkaido, Miyagi Prefectural Cancer Center north of Tokyo, Aichi Cancer Center near Nagoya, the National Kure Hospital near

Hiroshima, the National Shikoku Cancer Center in the island of Shikoku, and the National Kyushu Cancer Center in the southern island of Kyushu.

Every week, a lecture-style teleconference is held between the centres to exchange information on diagnostic techniques and treatments for cancer.

About a dozen other medical research institutes, medical associations and hospitals have joined the network with lower capacity links (mainly 64 kilobits per second) allowing them to access databases at the cancer centre. And several more will join next month. In collaboration with Nippon Telephone and Telegraph (NTT) and other telecommunications companies, the National Cancer Centre now has higher capacity 45-megabit and 156-megabit-per-second links between its two campuses.

There are also high-capacity links to the National Cardiovascular Centre in Osaka and the National Institute of Radiological Sciences in Chiba. The transmission of super-high-definition images is now being tested in collaboration with NTT.

Although the cancer centre's telemedicine network is widely used for clinical work, it is still funded as a research project by the Ministry of Health and Welfare on the grounds that it remains 'experimental' and does not come under the national health system.

But the ministry is planning to link 250 hospitals and medical organizations throughout Japan in March in a practical application of networking in medicine. This network, HOSPnet, will have an annual budget of ¥3 billion (US\$24 million). It will initially have only fairly low-capacity links and most hospitals will use an Internet telephone system for communication rather than teleconferencing. But the ministry has ambitions to upgrade it to a telemedicine network.

David Swinbanks

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Precious package: the \$100-million NICMOS will boost the scientific capabilities of Hubble.

In-flight service will let Hubble see red

[WASHINGTON] The Hubble Space Telescope will have its vision extended into infrared wavelengths next week when spacewalking astronauts install two new science instruments in the orbiting observatory and refurbish its tape recorders, guidance sensors and other hardware.

"We're replacing 1970s technology with 1990s technology," says Edward Weiler, project scientist for the telescope at the US National Aeronautics and Space Administration (NASA). Although there are risks in tinkering with a spacecraft that is operating perfectly, the payoff will be a "dramatic increase in science throughput", he promises.

One of the new instruments, the \$100-million Near Infrared Camera and Multi-Object Spectrometer (NICMOS), will more than double Hubble's spectral coverage, which at present observes only at visible and ultraviolet wavelengths.

The instrument, a combined camera, spectrometer, polarimeter and coronagraph, will be sensitive in the near infrared range from 0.8 to 2.5 micrometres. Astronomers expect to use it to see some of the youngest objects in the early Universe.

On the fourth day of the space shuttle Discovery's nine-day mission, astronauts will remove two first-generation spectrographs from the telescope and replace them with NICMOS and another new science instrument, the Space Telescope Imaging Spectrograph. This device takes many spectra at once, and so can work 30 to 40 times faster than the instrument it replaces. It will be especially useful for measuring the velocities of material spiralling into black holes.

The instruments will be checked and calibrated after installation, and will not be fully operational for several months. They are already much in demand: NASA received 1,400 proposals from astronomers for the next Hubble viewing 'cycle' — more than for any previous round.

Tony Reichhardt

Top FDA official named as 'de facto' chief

[WASHINGTON] Michael Friedman, a former US National Cancer Institute official, will become 'de facto' head of the Food and Drug Administration (FDA) later this month, when commissioner David Kessler departs, possibly to become dean at Yale University School of Medicine.

Friedman is currently deputy commissioner for operations at the FDA. His elevation does not indicate that he heads a list of candidates for commissioner, says a spokesman for Donna Shalala, the Health

and Human Services Secretary. Shalala is conducting a wide search for candidates, and soliciting "lots" of input from industry, consumer groups, Congress and the medical and scientific communities.

Neither the FDA nor Yale's School of Medicine would comment on reports that Kessler is a leading candidate to become dean of the school. A Yale spokeswoman says Kessler has visited the campus at New Haven, Connecticut, but that "nothing has been settled".

Meredith Wadman

US urged to monitor some genetic tests

[WASHINGTON] Declaring that "the public is not being adequately protected", an expert panel last week recommended several broad measures expanding government regulation of genetic testing. These include the establishment of a National Genetics Board and the monitoring of genetic tests considered to require "stringent scrutiny" by the Food and Drug Administration (FDA).

The panel also proposed that commercial laboratories be required to meet upgraded standards for genetic testing, or risk losing reimbursement by third-party payers, including huge government health programmes.

The recommendations, published last week in the *Federal Register*, come from the 15-member task force on genetic testing set up by the Working Group on the Ethical, Legal and Social Implications of Human Genome Research, established by the National Institutes of Health (NIH) and the Department of Energy.

The task force was established in 1995 to recommend steps to ensure the development and appropriate use of safe and effective genetic tests. A central recommendation is that some tests require "stringent scrutiny" throughout their development and marketing.

The suggested criteria used to decide whether stringent scrutiny is necessary would include: a test's potential to predict serious future disease in healthy people and their offspring; the absence of a confirmatory test; a test's uncertainty (the chances that a person with a positive test will not develop the disease); and the safety and effectiveness of known clinical interventions.

These criteria seem aimed in particular at tests being marketed by companies such as Myriad Genetics of Salt Lake City, Utah, which last October introduced full sequence commercial testing of *BRCA1* and *BRCA2* genes, mutations of which predispose women to breast and ovarian cancer.

Tests requiring stringent scrutiny would be subject to more careful review at several stages. They would be reviewed more carefully by institutional review boards, the local committees charged with ensuring that human subjects receive protection in experimental protocols.

They would also require pre-market approval by the FDA, whether sold as kits or carried out in-house. In-house tests are not at present monitored by the FDA, and virtually all genetic tests — including Myriad's — are being developed and marketed as such.

Separately, the task force recommends that Donna Shalala, the Health and Human Services Secretary, should create within her office an Advisory Committee on Genetics and Public Policy to develop "coordinated and consistent" genetic testing policies

throughout the department. That would include the NIH, the Centers for Disease Control and Prevention and the FDA.

This advisory committee would in turn recommend that Shalala establish a National Genetics Board of consumers, physicians, insurers and scientists. Its duties would include fully developing the 'stringent scrutiny' standard, and possibly reviewing and approving all experimental protocols for tests requiring such scrutiny, as a condition of their receiving federal funding.

If the FDA were to regulate in-house tests, then non-federally funded test developers would also be obliged, under current FDA regulations, to have protocols requiring stringent scrutiny approved by local institutional review boards, the task force says.

In the area of laboratory quality, the recommendations would "address the deficiencies" in monitoring genetic tests of the 1988 Clinical Laboratory Improvement Amendments — the law that established current quality standards for commercial laboratories.

The task force calls for the setting-up under that law of a national accreditation programme to ensure genetic testing quality, with third-party payers encouraged to reimburse only accredited laboratories.

Patricia Barr, a lawyer and task force member who also represents the National Breast Cancer Coalition, says that the recommendations will, if accepted, ensure "a level of coordination and regulation of genetic testing that is not taking place today".

Wayne Grody, a task force member who is a medical geneticist at the University of California, Los Angeles, says: "The most important recommendation is that there be some overarching body within Health and Human Services that would monitor and advise on genetic testing."

But the recommendations — particularly the proposed National Genetics Board and the call for FDA monitoring — are already drawing fire from academic and industry critics. Academics fear that the creation of a National Genetics Board could slow down the grants process.

Francis Collins, director of the National Human Genome Research Institute, declined to comment last week, but voiced concern during a task force meeting last December. "If this national board is slow to review and respond to protocols you may hear a hue and a cry and justifiably so," he said then.

Others say that involving the FDA could deal a fatal blow to small companies that develop tests. "It's going to take genetics back to the dark ages" if companies marketing in-house tests have to go through the FDA, says one industry critic. Only large companies with deep pockets could afford to develop tests in such circumstances, says the critic.

Comments on the panel's recommendations may be e-mailed by 10 March to tfgt-a@welchlink.welch.jhu.edu, and will then be finalized. Comments will be discussed at a meeting on 17–18 March in Baltimore, Maryland, after which final recommendations will be issued. **Meredith Wadman**

Russian budget falls short of 'legal level'

[MOSCOW] The State Duma — the lower chamber of the Russian parliament — last week adopted a budget for 1997 allocating 15.4 thousand billion roubles (about US\$2.5 billion), representing 2.9 per cent of all state spending, to research and development.

This contrasts with a commitment to give research at least four per cent of state spending contained in the so-called 'science law' signed by President Boris Yeltsin last August. The difference between the two figures is attributed to a lack of coordination between the executive and legislative branches of government.

"The Duma's law was prepared at the time when the government had already fixed all the major budget parameters," says Mikhail Glubokovsky, deputy chairman of the Duma's committee on science and education.

Glubokovsky's committee is keen to see the immediate implementation of a Duma resolution that the state should repay to research institutions the sum of 2.7 thousand billion roubles (US\$450 million) which

remains outstanding from last year's budget.

Vladimir Fortov, chairman of the state committee on science and technology, said that "the Russian science ship is in low water



Fortov: bemoans low science spend.

since the state spends on it only 0.3 per cent of gross national product".

Fortov, who is also vice-president of the Russian Academy of Sciences, said that this was because,

although it had been planned to allocate 0.5 per cent of gross national product to research, only 60 per cent of this had been paid. He added that financing is taking place on a week-to-week basis, preventing plans from being made "even a month ahead".

Speaking at a meeting of the scientific council of the Joint International Nuclear Institute in Dubna, Fortov expressed the hope that some of the international credits being received by Russia should be used to support a few large high-technology projects. **Carl Levitin**

German policy paper 'focuses on economic returns from space'

[MUNICH] Germany's research minister, Jürgen Rüttgers, is reported to have prepared a policy document, due to be presented to the German government this month, setting out ways in which federal funding of space research might be focused more tightly on projects that promise an economic return. The proposal includes the merger of the German Space Agency (DARA) with the national air and space research centre (DLR).

According to the newspaper *Handelsblatt*, the policy document criticizes the German space industry for failing to exploit technology such as telecommunications satellites, and for an excessive focus on building "innovating prototypes and other unique systems".

Rüttgers also claims that "substantial deficits" exist in the activities of the European Space Agency. Rüttgers says that the agency's management costs are too high, and that Germany should be allowed a greater role in its decision-making process, in line with the size of its financial contribution. But Rüttgers is also said to have confirmed that participation in the international space station should remain at the heart of Germany's space activities.

Risk panel recommends better communication

[WASHINGTON] US government agencies were recommended last week to adopt more sophisticated methods for estimating and communicating health and environmental risks. The 10-member Commission on Risk Assessment and Risk Management, jointly appointed in 1994 by Congress, President Bill Clinton and the National Academy of Sciences, encouraged regulators to include multiple factors in their risk assessments, rather than focusing on individual chemicals or single environmental media. The panel also favoured consensual regulatory action over increased judicial review, and said that economic analysis, while useful, should not be the "sole or overriding determinant" of regulatory decisions. The commission's report can be found on the World Wide Web at www.riskworld.com.

Nuclear waste expert criticizes Nirex

[LONDON] Sir John Knill, a former chairman of the UK government's Radioactive Waste Management Advisory Committee, has criticized the waste disposal company Nirex for refusing to acknowledge reservations about the adequacy of its data at Sellafield in Cumbria, in northwest England, the

proposed site for a nuclear waste repository.

In a letter to John Gummer, the UK environment secretary, Knill wrote that during discussions with the committee the company had "consistently denied" having reservations over its safety assessment programme, as well as the existence of differences between hydrogeologists advising Nirex and the company's own modellers. The letter appears to have been prompted by the leaking of an internal memorandum written by Nirex's director of science, voicing concern at the lack of datapoints at the Sellafield site (see *Nature* 385, 282; 1997).

Parents file lawsuit over *E. coli* death in Japan

[TOKYO] The parents of a 12-year-old girl who died last August after eating a school lunch infected with a highly toxic strain of *Escherichia coli* have filed a lawsuit against their local government near Osaka in Japan claiming negligence and demanding ¥78 million (\$680,000) in compensation. This is the first of several lawsuits expected to be filed by the families of victims of last summer's *E. coli* infection which affected about 10,000 people, most of them schoolchildren.

The suit claims that the Sakai municipal government could have foreseen the danger

To **save an hour**
each time you
purify histidine-
tagged proteins,
put it on the
tip of a syringe



Darwin a better name than Wallace?

Sir — C. Ambrogio Lorenzini (*Nature* 384, 508; 1996) highlights the possibility that double-barrelled names may be incorrectly listed in publication databases. This is an important issue because citation rates are increasingly used to assess the value of research. But it is not only those with unusual names who may get a raw deal out of citation indices. Investigation of the relationship between alphabetical position of surname and citation rate reveals that researchers nearer to Darwin may appear to be making a bigger contribution than those nearer Wallace.

The number of papers published by authors of a particular surname initial for 1994 was taken from the on-line *Science Citation Index (SCI)*. Citation rates by initial are not easily gleaned from this source, so I measured the number of column centimetres of citations in the paper version of the 1994 *SCI*. Comparison of these data reveals a clear decrease in citations per publication with surname initial (linear regression $r^2 = 0.33$, $F = 12.04$, $P < 0.002$). An obvious

explanation is that authorship of papers is sometimes determined alphabetically. Citations are based only on first-author papers, whereas the on-line database provides information on all authors, so those with later initials may be penalized through being less likely to be first author.

This doesn't, however, appear to be sufficient explanation. By measuring column centimetres per author in the index of 10 randomly chosen 1994 editions of *Current Contents*, which lists only first authors, one is able to estimate publications by first authors only. This also reveals a significant decrease in number of citations with position in the alphabet (linear regression $r^2 = 0.33$, $F = 16.7$, $P < 0.0004$). Because the citation index goes back many years, this result could arise from a past tendency to put authors in alphabetical order. But using order of authorship in *Letters to Nature* from the first issue of the past 25 years fails to support this idea. Taking into account the increased number of authors on more recent papers, there is no decrease

in the proportion of papers with alphabetical authors since 1972 (linear regression $P = 0.17$).

So what is the explanation? Could it be that researchers tend to flick through reference lists and to have had their fill before they reach the end? Do we go to the library with an alphabetical list of papers to read and run out of steam half-way through? Certainly the effect is very robust; removing the five least cited initials from either regression does not drop significance below 0.05, so clearly something is going on.

I suggest that, whatever the explanation, in fairness to those relegated to the end of reference lists, a correction factor should be applied using the slope of the regression of citation rate versus author initial. By this method, the citation rate of someone whose initial is, say, T should be multiplied by 1.36 — quite a boost.

Tom Tregenza

Department of Biology,
University of Leeds,
Leeds LS2 9JT, UK

A novel paradigm

Sir — We wish to report a novel paradigm. In recent years, we have observed an increase in the number of scientists who use the word 'novel' to describe their work. This observation has been borne out by a statistical analysis of the Medline database.

The figure plots frequency of the word 'novel' in titles and abstracts as a proportion of total papers. This shows the trend to greater novelty over time; in fact, the change is well fitted by an exponential. Two possible interpretations are (1) novelty in research is actually increasing over time at an exponential rate, thereby leading to the exciting possibility that in the morning hours of 7 February 2020 all science will be novel; or (2) the probability of someone using the word 'novel' to describe their

work is proportional to the probability that they have seen this word used in a similar context before. This in turn is roughly proportional to the number of recent appearances in the literature.

If the latter mechanism is at work, the inclusion of 'exposure saturation' and 'staleness' factors in a mathematical model should lead to predictions of a novelty maximum and possibly a novelty decline. It is of interest to note that 'paradigms' are increasing in appearance exponentially as well. Indeed, the future looks very bright.

Simon H. Friedman

Department of Pharmaceutical Chemistry,
University of California,
San Francisco, California 94143-0446, USA

Jens O. M. Karlsson

Department of Mechanical Engineering,
University of Illinois,
Chicago, Illinois 60607, USA

complementary medicine, published in different countries by different publishers: *Research in Complementary Medicine*, *Complementary Therapies in Medicine*, *Alternative Therapies in Health and Medicine* and *Natura Med.* (The editor of a fifth journal, *Alternative and Complementary Medicine*, did not wish to participate in our survey.)

We categorized all 204 articles in one year's editions of these journals as positive (a particular intervention is helpful for a particular condition), neutral (no clear conclusion) or negative (intervention is unhelpful). The pattern in all four journals was strikingly similar, with 64% of papers classified as helpful, 35% neutral and only 1% negative.

Although our results are limited because our survey is small, they demonstrate a strong publication bias in favour of positive conclusions about alternative therapies. If confirmed, our findings imply that the literature in this field is not objective.

E. Ernst

M. H. Pittler

Department of Complementary Medicine,
University of Exeter,
Exeter EX2 4NT, UK

Alternative therapy bias

Sir — Complementary medicine is becoming increasingly popular: it is estimated that between a quarter and a half of the population in the United States, Europe and Australia have used this type of therapy. But how credible is this field?

We have analysed the conclusions of papers in four leading journals of

corres@nature.com Letters submitted for Correspondence may be e-mailed to corres@nature.com. Do not send items intended for any other section by e-mail unless requested to do so.

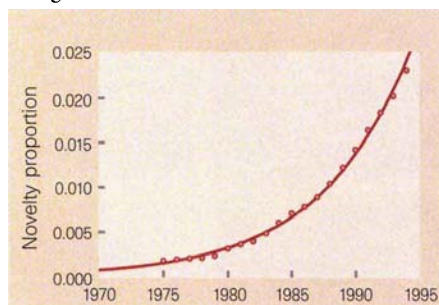


Figure 1 Proportion of papers per year in the Medline database that contain the word 'novel' in the title or abstract. ($y = 7.3749e - 126e^{(0.14263x)}$, $R = 0.99811$)

How do memories leave their mark?

Mark F. Bear

We remember some events while others slip from our minds. The results of a new study indicate how records of sensory experiences might be consolidated into long-term memory.

Sensory experiences leave their mark on the brain by altering the effectiveness of synapses between neurons. Based on how active they are during a sensory experience, some synapses on a neuron grow stronger, while others grow weaker; and the pattern of synaptic change represents a memory of the experience. Changes in synaptic strength are due to modifications of existing synaptic proteins, by phosphorylation and dephosphorylation, for example. However, unless these modifications are consolidated, the synaptic strengths decay back to their original values and the memory is lost.

Consolidation requires protein synthesis¹. Presumably, the newly made proteins are transported to the modified synapses where they make the temporary changes permanent. But how do they find the correct synaptic address? The results of a clever new study have led Frey and Morris² (on page 533) to suggest an answer.

Frey and Morris used a popular synaptic model for memory: hippocampal long-term potentiation (LTP). LTP is a prolonged increase in the strength of synapses that have been stimulated repetitively at high frequencies (for example, 100 Hz; Fig. 1a). The duration of the change depends on how much of this rapidly repeated ('tetanic') stimulation is given. A single, brief tetanus produces a large increase in synaptic

strength, but this decays back to baseline values within a few hours (Fig. 1b, single upward arrow). Repeated, longer tetani produce about the same change in synaptic strength, but the change persists for longer than 8 hours (Fig. 1b, three upward arrows). LTP that lasts for longer than 3 hours is termed late LTP.

The main difference between single and repeated tetani is revealed by the effect of agents that inhibit protein synthesis. Such agents lead to a reduction in the time course of LTP that is caused by repeated tetani, so that it resembles the decaying LTP that is caused by a single, brief tetanus (Fig. 1b, dashed line). Along with other observations, this has led to the idea that not only do repeated tetani produce post-translational modifications in the existing synaptic proteins that cause the initial potentiation, but they also trigger a wave of protein synthesis that causes the LTP to become consolidated³.

Frey and Morris set out to determine whether the newly made proteins would act only on those synapses whose stimulation had triggered their synthesis. To do this, they took advantage of another property of LTP called 'input specificity', which means that potentiation occurs only at (or near⁴) synapses that are stimulated during the tetanus. So if a tetanus was delivered to the first synapse in Fig. 1, only that synapse

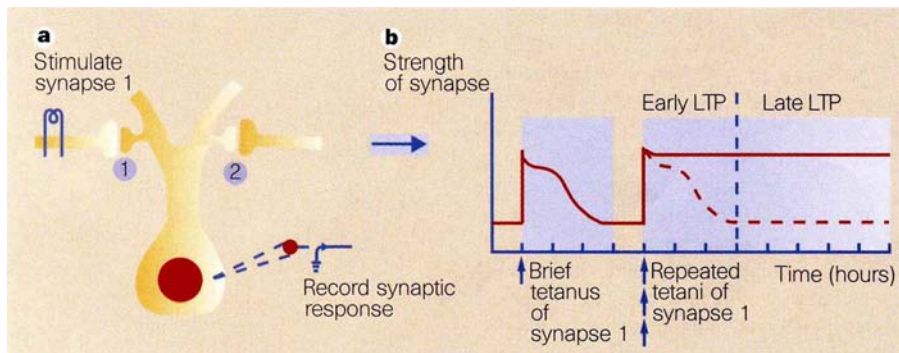


Figure 1 The memory of a sensory experience is formed by a pattern of changes in the strength of synapses. **a**, Individual synapses can be stimulated repetitively at high frequency ('tetanic stimulation'). **b**, The resulting increase in the strength of the synapse can be recorded. A brief tetanus (single upward arrow) produces LTP which lasts about 3 hours. Repeated tetani (three upward arrows) produce a change in synaptic strength that lasts longer than 8 hours. If a protein-synthesis inhibitor is added during, and immediately after, the repeated tetanization, the duration of LTP is much less (dashed line). The protein-synthesis-dependent LTP that lasts longer than 3 hours is known as late LTP.

would be potentiated — the second synapse would be unchanged.

The idea behind the experiment was to stimulate the first synapse with repeated tetani to trigger the wave of protein synthesis, and then ask if these proteins would consolidate LTP induced by a single tetanus at the second synapse some time later (Fig. 2, overleaf). They found that the wave of protein synthesis that is triggered by repeated tetani at the first synapse does lead to consolidation of LTP at the second synapse — but only if the single tetanus to the second synapse is delivered within 90 minutes of the initial repeated tetanization. So the new proteins are not targeted specifically to the synapses whose activity triggered their synthesis; they are available to all the synapses on the neuron. However, only the synapses that are tetanized within the 90-minute time-window can use this new resource.

This work has several implications. The involvement of gene expression in establishing long-term memory had suggested that memories may actually reside in the nucleus, and that they are maintained by self-sustaining changes in gene transcription. But the new results indicate that all that may be necessary is a transient change in gene expression, to provide a resource that can prolong an otherwise temporary change in synaptic strength⁵. Another implication is that the induction of LTP creates a synapse-specific receptor — or tag — that accepts the resource provided by gene expression. If the resource is available and the tag is present, the synaptic enhancement is consolidated.

Two obvious questions remain: what is the resource, and what is the synapse-specific tag? The resource, which could be one or several proteins, is likely to be the product of immediate-early gene expression, which is under the control of the cyclic-AMP-response element (CRE).

The reasons to suspect the involvement of immediate-early genes are the rapid onset and brief duration of the essential wave of protein synthesis, and the fact that these genes are turned on by strong synaptic stimulation⁶. There are several reasons to suspect that expression of the resource is controlled by CRE: CRE-mediated gene expression occurs in hippocampal neurons following repeated, but not single, tetani⁷; and disruption of CRE regulation prevents the induction of late LTP with repeated tetani⁸.

A good guess as to the identity of the synapse-specific tag would be a synaptic phosphoprotein. One intriguing possibility is that the tag is a substrate for cAMP-dependent protein kinase, as there is evidence that cAMP is generated at tetanized synapses⁹. This could explain why the addition of cell-permeable analogues of cAMP to hippocampal slices produces a nonspecific

synaptic strengthening that resembles late LTP¹⁰. Treatment with cAMP could therefore potentially both tag the synapses and stimulate production of the resource.

The story holds together nicely for the LTP model, but what about real-life memories? One of the most exciting developments in neurobiology over the past few

years has been the discovery that CRE-mediated gene expression is required for the establishment of long-term memory in species ranging from fruitflies to mice¹¹. The new results indicate that these CRE-dependent proteins will consolidate (into long-term memory) events over and above those that triggered the gene expression in the first place. Where were you when Kennedy was shot? If you remember, Frey and Morris would suggest that the synapses in your brain that were transiently tagged by the experience of your location were consolidated by the wave of protein synthesis that was triggered by the news of the assassination.

Because the consolidation process is not restricted to the synapse by which it was triggered, it could be subject to modulation by other signals. Interestingly, the profound amnesia that is produced by lesions of the medial temporal lobes is qualitatively similar to that produced by protein-synthesis inhibitors. Perhaps the medial temporal lobes exert their effects on memory consoli-

dation by chemically modulating gene expression in neurons throughout the cerebral cortex. Understanding such modulation could open new avenues for the treatment of the memory loss that often accompanies traumatic brain injury, stroke, Alzheimer's disease and normal ageing. □

Mark F. Bear is in the Howard Hughes Medical Institute, Department of Neuroscience, Brown University, Providence, Rhode Island 02912, USA.

1. Squire, L. R. *Memory and Brain* (Oxford Univ. Press, New York, 1987).
2. Frey, U. & Morris, R. G. M. *Nature* **385**, 533–536 (1997).
3. Nguyen, P. V., Abel, T. & Kandel, E. R. *Science* **265**, 1104–1107 (1994).
4. Schuman, E. M. & Madison, D. V. *Science* **263**, 532–536 (1994).
5. Lisman, J. J. *NIH Res. J.* **7**, 43–46 (1995).
6. Abraham, W. C., Dragunow, M. & Tate, W. P. *Mol. Neurobiol.* **5**, 297–314 (1991).
7. Impey, S. et al. *Neuron* **16**, 973–982 (1996).
8. Bourchouladze, R. et al. *Cell* **79**, 59–68 (1994).
9. Chetkovich, D. M. & Sweatt, J. D. *J. Neurochem.* **61**, 1933–1942 (1993).
10. Frey, U., Huang, Y. Y. & Kandel, E. R. *Science* **260**, 1661–1664 (1993).
11. DeZazzo, J. & Tully, T. *Trends Neurosci.* **18**, 212–218 (1995).

Quantum optics

The amazing atom laser

Keith Burnett

The word laser now has a strong emotional impact, conjuring up images of space-battles and high-tech surgery. There was a time when this strange acronym, as yet unloved by science-fiction writers, conjured up nothing at all to a general audience. What is more, scientists thought of the laser as nothing more than a curiosity brought forth by optical physicists and engineers, fascinating perhaps, but little more. It was termed, admittedly by the unkind, 'a solution in search of a problem'.

We now have the beginnings of a new type of solution — a new type of laser. This is not a source of light but of matter, in the form of alkali atoms. In the 27 January issue of *Physical Review Letters*¹, a group at MIT led by Wolfgang Ketterle describe an 'output coupler' that allows pulses of matter to escape in a narrow beam from a trapped Bose–Einstein condensate (Figs 1 and 2). In the 31 January edition of *Science*², the same group show that this beam is coherent, an essential test of laser behaviour.

The possibility of an atom laser should not be such a shock if we recall that all matter has a wave-like nature, just like light, with each particle's de Broglie wavelength determined by its momentum. So if light is made of waves and can be produced in laser form, why not a laser-like source of matter waves?

The problem is how to do it. In the case of light, we rely on a very important property of photons. Energy is first put into a material form — the excited states of atoms or molec-

ules — and then re-emitted in the form of photon energy packets; lasers are possible because photons can be created with exactly the same direction and speed as those that are already travelling through the medium. This means that the number of photons travelling through a laser medium can become enormously amplified. This property was elucidated first by Einstein and Bose in their studies of the implications of wave mechanics for the theory of radiation.

So can this idea be extended from photons to other particles, such as atoms?

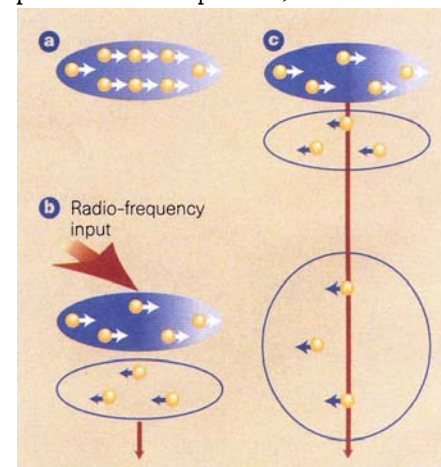


Figure 1 Radio leakage. The magnetically trapped Bose–Einstein condensate (a) is separated into two states with opposing spins by a radio-frequency input (b). The state with flipped spin is no longer held by the magnetic trap, and so falls out (c).

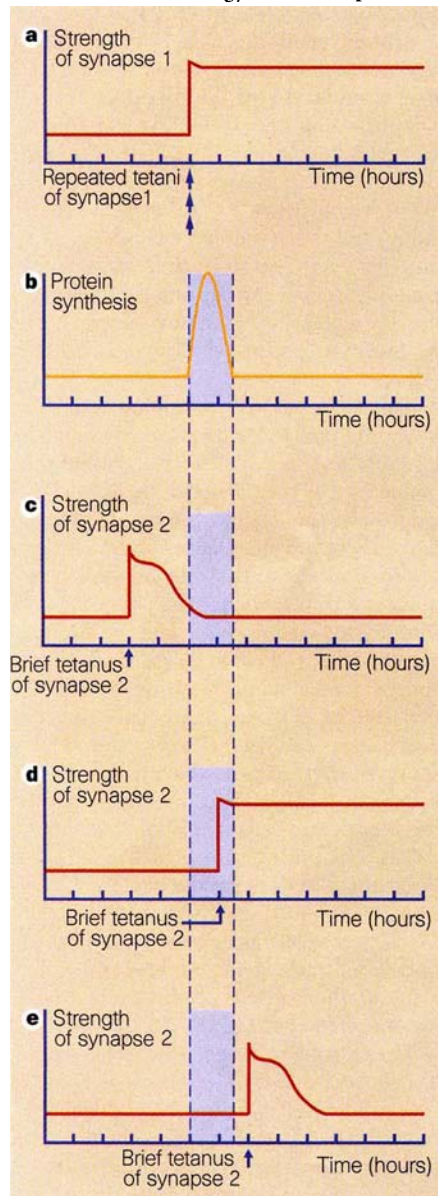


Figure 2 Measurements made by Frey and Morris² to determine whether the proteins induced by repeated tetanization of synapse 1 allow consolidation of a single tetanus at synapse 2. a, Repeated tetani are delivered to synapse 1 to trigger a wave of protein synthesis. b, Time course of the presumptive increase in protein synthesis required for consolidation of LTP. c, Effect on synapse 2 of a single tetanus delivered before the wave of protein synthesis is triggered; no late LTP is produced. d, Effect of a single tetanus delivered to synapse 2 during the wave of protein synthesis; late LTP is produced. e, Effect of a single tetanus at synapse 2 after the wave of protein synthesis; no late LTP is produced.

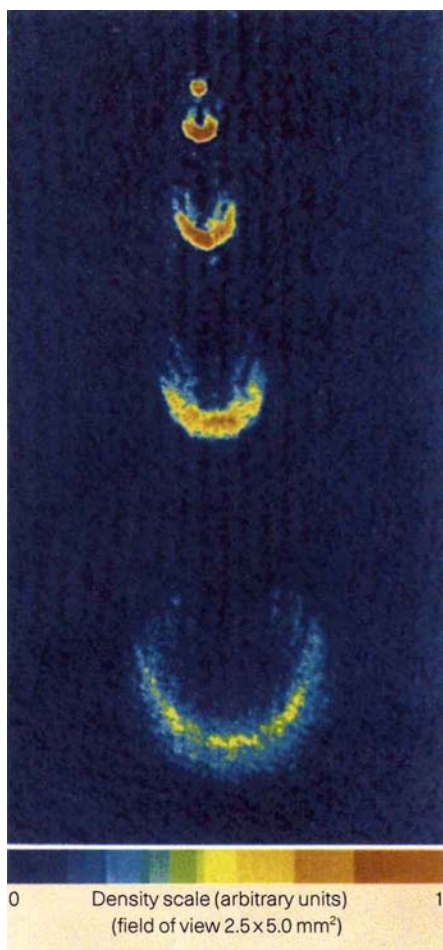


Figure 2 The atom laser in action. Drops of coherent sodium fall under gravity, about five milliseconds apart.

Indeed it can, if, like photons, the particles are Bose–Einstein particles (bosons for short). All particles are either bosons, with an integral number of spin units, or fermions, with half-integral spin. The crucial physical difference between the two is that no two fermions can occupy the same quantum state, so coherent behaviour is impossible.

Most alkali metals are bosons, so a bunch of alkali atoms can join together in identical states of motion. At sufficiently low temperatures, Einstein and Bose showed that the propensity to do this sets in with a vengeance, and all the atoms condense into the lowest-energy state — hence the term Bose–Einstein condensation. It is done in practice by cooling a magnetically trapped gas of atoms to extraordinarily low temperatures, below a microkelvin. So far, no other way has been found to make a pure laser-like assembly of atoms. A Bose–Einstein condensate was first made using rubidium in 1995, at the Joint Institute for Laboratory Astrophysics in Colorado³, and later that year at MIT.

In the new experiments at MIT, they have made the next step in producing a laser-source of atoms: the smooth transference from the magnetic box in which the atoms

are corralled to an atomic beam¹ (Fig. 1). At present, the beam is simply collimated by the magnetic fields of the trap, but in principle it can be directed at will (for animations and other details see <http://bink.mit.edu/dallin/news.html#matterwave>). Ketterle and colleagues have also shown how to test the properties of the beam very precisely². They made a pair of sources, shining both of them on a screen, and looking at the regions where they overlap. If two such beams are coherent (that is, pure enough in wavelength) then a regular pattern of stripes will be observed on the screen, just as the MIT group saw.

These experiments have produced exquisite results that can go directly into the textbooks. There are other aspects of laser light, such as continuous output and stable intensity, that cannot yet be reproduced in the matter beam, and many would want to reserve the term atom-laser for the next generation of sources. But what cannot be doubted is that a huge step has been made towards making laser-like sources of atomic de Broglie waves a tool for science and technology.

The new source shares many of the practical advantages of laser light. It is an intense, precisely directed, beam of matter. It should be possible to focus such a beam using diffraction gratings or the equivalent of lenses for atomic waves — an atom passing through a laser beam that has an intensity gradient feels a position-dependent force, so a beam with the right geometry acts as a lens. Therefore, atoms can be directed to a point or feature on a material surface, for example. This is precisely what we need to produce systems on the nanometre scale, such as micro- (or rather nano-) electronic circuits.

Coherent matter waves may also surpass laser light beams as fundamental measuring sticks. The spacing between the lines on the ruler is set by the beam's wavelength, which is around a half a micrometre for visible light. But with matter waves, the wavelength can be tuned by changing their speed (light, of course, moves with a fixed speed in a given medium) and one should be able to produce matter waves with wavelengths a hundred times smaller than that of a visible photon. Accurate, continuous-wave atomic clocks will permit a wide range of new measurements. Atom-laser interferometers may be the best gyroscopes and accelerometers, making more precise measurements of gravity than ever before. They will also be used to test certain quantum theories that predict a divergence from conventional quantum behaviour on large scales. □

Keith Burnett is in the Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, UK.

1. Mewes, M.-O. *et al.* *Phys. Rev. Lett.* **78**, 582–585 (1997).

2. Andrews, M. R. *et al.* *Science* **275**, 637–641 (1997).

3. Anderson, M. H., Ensher, J. R., Matthews, M. R., Wieman, C. E. & Cornell, E. A. *Science* **269**, 198–201 (1995).



100 YEARS AGO

The current number of the *Comptes rendus* contains a description of an absolute electrometer intended for measuring small electromotive forces (about 1 volt), designed by MM. Pérot et Fabry. The instrument consists of an attracted disc electrometer, in which the necessary sensitiveness is obtained by greatly reducing the distance between the plates. The attracting disc consists of the plane end of a glass cylinder about 6 cm. in diameter and 1 cm. high, this height being large compared with the distance between the two discs.... These discs are lightly silvered, and their parallelism adjusted, and their distance measured by being traversed normally by a beam of monochromatic light, which forms interference bands between the light which has passed directly through the thin silver coating and that which has been reflected an even number of times at the silvered surfaces. The electrical attraction is measured by comparing the deformation produced in three springs, which carry the movable disc by the electrical forces, with that produced by a known weight when placed on the movable disc. The authors have obtained 0.0048467 as the mean value of the electromotive force of a Clark cell at 0° in electrostatic measure, and, taking the electromotive force in electromagnetic units as 1.4535×10^8 , the value of v , the ratio of the units obtained is $v = 2.9989 \times 10^{10}$.

From *Nature* 4 February 1897.

50 YEARS AGO

In studying photographic plates exposed to cosmic rays, we have found a number of multiple disintegrations each of which appears to have been produced by the entry of a slow charged particle into a nucleus.... In view of the evidence provided by [our] experiments, we believe that the particles producing the observed disintegrations are negative mesons. We consider our photographs as evidence of the fact that the entry of the particle into the nucleus produces excitation, which is followed by the emission of heavy particles. In support of the general picture of the fate of the positive and negative mesons, we find examples of tracks which end in the emulsion and which we identify as positive mesons.

G. P. S. Occhialini and C. F. Powell

From *Nature* 8 February 1947.

Diamonds from warm water

R. C. DeVries

There is a revival of interest in growing diamonds hydrothermally in the laboratory, but it has met with questionable success. One dream is to grow large diamond crystals in mild conditions, at about 500 °C and 1 kbar (a thousand atmospheres), in the manner so successful with quartz and emerald. A more reasonable expectation may be the extension of our knowledge about how natural diamonds grow. On page 513 of this issue¹, Zhao and colleagues present the latest efforts to grow diamond under relatively low pressure and temperature, using a solution of nickel and carbon in water.

Carbon can be made into diamond directly by compression, both in industry and in nature (in meteorite impacts, for example). That requires a pressure above about 120 kbar and a temperature of about 2,000 °C. At lower pressure and temperature (50 kbar and 1,400 °C), indirect, static conversion via solutions of carbon in iron and/or nickel occurs on a large scale every day in factories all over the world, to produce small, abrasive diamond grains with metal inclusions. If diamond is still stable at core-of-the-Earth pressures (3.6 Mbar), and if there's enough carbon present in the iron down there, we might expect the same process to produce large diamond crystals (perhaps even Scott Fitzgerald's "Diamond as big as the Ritz"). But there is a third route, the hydrothermal process, which may be responsible for most natural diamonds.

Most diamonds that we know about have come into the crust along paths that tapped at least the upper mantle, and perhaps the deeper mantle too. There are still many mysteries in the genesis of natural diamonds, but it is clear that they are remarkably free of metallic elements (that is one way to differentiate synthesized from natural diamond gemstones). So metal-carbon solutions do not appear to be involved in diamond growth in the mantle and above.

Other inclusions in natural diamonds (such as silicates and sulphides) are a prolific source of information about the growth environment. But more pertinent to hydrothermal synthesis is the clear evidence for carbon, hydrogen and oxygen in the form of various liquid phases of the C-H-O system. The presence of oxygen and hydrogen in diamond has been known for a long time², but characterization of 'coated diamond' — a poor-quality, fibrous diamond layer overgrown on a more homogeneous base — strongly supports the presence of liquid C-H-O phases during diamond growth³. The chemistry of graphite in equilibrium

with water in this system has been clearly established⁴, and it is a reasonable assumption that, with increased pressure, diamond would form from the same chemistry. The very small diamonds found in metamorphic rocks⁵ probably grew from similar C-H-O liquids.

It is pertinent that the major breakthrough in diamond synthesis after the early high-pressure successes was chemical vapour deposition of diamond⁶ at less than one atmosphere and 1,000 °C, in the C-H and C-H-O systems. It can also be grown from metal-C-H systems at similar pressure and temperature⁷. The mechanism depends on the presence of atomic hydrogen, which is probably not involved in hydrothermal growth, but the idea of a broad continuum of diamond stability is still valid. Diamond deposition also occurs in molten NaOH in the presence of nickel and carbon^{7,8}, and carbon with a high level of *sp*³ bonding (as in diamond) can be deposited by the reaction of Al₄C₃ and CCl₄ in a molten halide melt at a pressure of just one atmosphere⁹. The problem is how to capitalize on these clues.

Zhao *et al.*¹ appear to have achieved diamond-on-diamond deposition under hydrothermal conditions in the presence of nickel and vitreous carbon, at 800 °C and 1.4 kbar. Their problem is proving that there is new growth of diamond in the presence of diamond seeds. The same difficulty was faced by the early workers in chemical vapour deposition, and there are still sceptics who question whether any success was achieved in that technique until the Russians grew crystals visible to the eye.

A common problem in this field is diamond fever, a disease that enables one to see diamonds where they are not. Many have succumbed to this affliction. But the authors are aware of the

problem and have brought all their resources (electron microscopy, energy-dispersive X-rays, Raman scattering, and their intuition) to bear on the grains they have produced. They conclude that "...aggregates, tens of micrometres in size, of small diamond crystals can be grown in a hydrothermal environment from a mixture of carbon, water and metal...".

They are reasonably sure that there were no big 'sports' in their original seed distribution, and that the anhedral 15-micrometre particle in Fig. 1a on page 513 is new diamond. If this grain grew from a sub-micrometre seed, the same growth rate would produce the Cullinan diamond in about 30 million years; compare that with a time of only about 200 days extrapolated from currently available growth-rate data for gem-quality diamonds grown by the thermal-gradient method from metal-carbon solutions. But even if the aggregates of Fig. 1b are sintered from the original seeds, the result is interesting. If reproduced by others, it will confirm that hydrothermal diamond growth is possible, and so may be involved in forming natural diamonds. That could explain why natural diamonds contain no metal.

There are unanswered problems, like the destiny of the nickel and its role in the process, and how to find a reliable method to distinguish seeds from new growth. Be sceptical, but this result looks like the state-of-the-art from a hydrothermal system. □

R. C. DeVries is at 17 Van Vorst Drive, Burnt Hills, New York 12027-9712, USA.

1. Zhao, X.-Z., Roy, R., Cherian, K. A. & Badzian, A. *Nature* **385**, 513–515 (1997).
2. Melton, C. E. & Giardini, A. A. *Am. Mineral.* **59**, 775–782 (1974).
3. Navon, O. *et al.* *Nature* **335**, 784–789 (1988).
4. Rumble, D. & Hoering, T. C. *Geochim. Cosmochim. Acta* **50**, 1239–1247 (1986).
5. Sobolev, N. V. & Shatsky, V. S. *Nature* **343**, 742–745 (1990).
6. Spitsyn, B. V. *et al.* *J. Crystal Growth* **2**, 219–226 (1981).
7. Cherian, K. A. *Surf. Coat. Technol.* **47**, 127–143 (1991).
8. Pavel, E. *et al.* *Diamond Rel. Mater.* **2**, 505–507 (1993).
9. French, E. A. & French, D. A. US patent no. 4,352,787 (1981).

Figure 1 Industrial diamond grit. These diamonds could have crystallized in the Earth's mantle from a watery solution.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Birds face sexual discrimination

Patricia Adair Gowaty

The crown jewel of evolutionary theory has long been thought to be R. A. Fisher's equilibrium sex-ratio theory¹. According to Fisher, when daughters and sons are of equal reproductive value, parents should adjust the production of their offspring in such a way that the average fitness costs of sons and daughters are equal. This means that if ecological or social factors — for example, the differential mortality of sons in many mammalian species — increase the fitness costs to the parents of one sex relative to the other, parents should discount these costs by producing fewer individuals of the more costly sex.

The study of birds has always been a tarnish on Fisher's theory because few studies have shown sex ratios that fit the predictions. But on page 522 of this issue, Komdeur and colleagues² present their observations of adaptive sex ratios among eggs of Seychelles warblers, *Acrocephalus sechellensis*. They report the largest skew in hatching sex ratios that have ever been seen in a bird species; and they show that these changes are consistent with the action of selection pressures on breeding pairs.

Until now, the relatively poor relationship between the predicted and observed adaptive sex ratios in birds has been

explained by the constraints on chromosomal sex determination that are imposed by meiosis: diploid meiosis 'guarantees' a 50:50 sex ratio of gametes. Furthermore, it is often difficult to assign sex to sexually monomorphic animals such as most young birds. In fact, fifteen years ago Charnov³ bemoaned the fact that nothing much of interest was going on with regard to progeny sex ratios in birds, probably because of such meiotic constraints and the difficulties of sexing young birds. In marked contrast are the extreme and precise sex ratios that are often reported for haplo-diploid insects such as parasitoid wasps and ants. So, given the theoretical adaptive consequences of variation in the sex ratios of offspring, and the remarkable correlation between the predicted and observed ratios in haplo-diploid insects, the apparent lack of convincing adaptive sex-ratio variation in wild birds was paradoxical.

Offspring can decrease their individual costs to parents through local-resource enhancement — for example, by helping their parents to raise later offspring^{4,5}. Conversely, they can increase their costs by competing with the parents for access to limited food supplies or other essential resources^{6,7}. Like differential mortality of one sex versus the other, local-resource enhancement and competition are potential sources of differential fitness effects of sons and daughters on parents; and they are probably the selection pressures that affect the sex ratios of offspring.

In an earlier study⁸, Komdeur showed that depending on the availability of food (insects) to Seychelles warblers, 'helpers' — adult daughters that remain on parental territories — decrease their individual costs by increasing their parents' reproductive success. They enhance the local resources by helping their parents to raise additional offspring; however, this only occurs on 'high-quality' territories, where food is plentiful. On low-quality territories, these helpers become 'hinderers': they increase their costs by decreasing their parents' reproductive success through competition for limited food.

The variation in the fitness costs of helpers allowed the sex ratios of offspring that would be produced by pairs without helpers to be predicted. In the new study, Komdeur *et al.*² determined the sex of day-old hatchlings by observing sex-specific bands from blood samples that had been subjected to 'randomly amplified polymorphic DNA' (RAPD) analysis. They found that Fisher's predictions are borne out: whereas pairs on high-quality territories produce 87% daughters (the sex that is retained as helpers), those on low-quality territories produce only 23% daughters. For a bird species, these sex ratios are extreme⁷: differences in the numbers of sons and daughters in birds are usually only of the order 2–3%, and occasionally as high as 10% (ref. 4).

The addition of data from experimental

Volcanology

Bubbling over

When a volcano blows its top, the ensuing display can be spectacular. Eruptive events such as 'fire fountains', shown here accompanying the November 1959 eruption of the Hawaiian volcano Kilauea, inspire thoughts of tremendous forces operating at depth in the Earth. But according to A. W. Woods and S. S. S. Cardoso, in a paper that appears in this issue (*Nature* 385, 518–520; 1997), underlying such phenomena is the fairly innocuous process of bubbles of gas separating out from molten rock.

The fuel for most volcanic eruptions is supplied by magma chambers — reservoirs of molten rock that can contain large quantities of dissolved gas. As the magma within such a chamber cools and starts to solidify, the dissolved gases are concentrated preferentially in the remaining melt. When the concentration of gas eventually reaches saturation point, bubbles form. For magmas rich in silica, the very act of bubble-formation can be sufficient to raise the pressure in the magma chamber to the point of eruption. But it is not an effective eruption trigger for basaltic magmas, as are found beneath the Hawaiian volcanoes.

Bubbles in basaltic magma might nevertheless have an important, albeit more indirect, role to play. Basaltic magmas are less viscous than their silica-rich counterparts: once formed, gas bubbles should be able to rise fairly readily through the magma and accumulate as a foam layer at the roof of the chamber. The ascent of the bubbles should in turn increase chamber pressure, because — put crudely — magma is an incompressible fluid and so the bubbles cannot increase in volume in

response to decompression. In consequence, the pressure within the rising bubbles does not change and the net effect is that pressure is 'transported' from deep in the magma to the top of the chamber.

Real volcanoes are more complex, however, because the competing processes of magma cooling (which increases the viscosity) and turbulent convection within the melt, as well as bubble ascent, have to be taken into account. Woods and Cardoso develop a model that encapsulates these processes in a self-consistent way, and they can reproduce — at least qualitatively — the eruptive behaviour of basaltic volcanoes such as Kilauea. It remains to be seen whether this model can be used as a basis for volcanic hazard assessment, but the first results are encouraging.

Karl Ziemelis

offer up. Second, what are the behavioural details of resource enhancement and competition under natural conditions? The solution lies in observations of the behavioural interactions between parents and their adult-aged offspring. Because of the possibility that adult offspring help by reducing the workloads of either the mother or the father, or that they hinder by decreasing the access of only one parent to essential resources, it is possible that different optimal sex ratios exist for mothers versus fathers. So understanding how parents resolve such parent–parent conflict¹⁰ over the sex ratio is another interesting question that studies on Seychelles warblers might illuminate. Finally, given the strength of the data obtained by Komdeur *et al.*, the apparently unexceptional empirical observations that have been

made on many other avian species beg the question, what makes Seychelles warblers so special? □

Patricia Adair Gowaty is in the Institute of Ecology, University of Georgia, Athens, Georgia 30602-2602, USA.

1. Fisher, R. A. *The Genetical Theory of Natural Selection* (Dover, New York, 1958).
2. Komdeur, J., Daan, S., Tinbergen, J. & Mateman, C. *Nature* **385**, 522–525 (1997).
3. Charnov, R. *The Theory of Sex Allocation* (Princeton Univ. Press, Princeton, 1982).
4. Gowaty, P. A. & Lennartz, M. R. *Am. Nat.* **126**, 347–353 (1985).
5. Emlen, S. T., Emlen, J. M. & Levin, S. A. *Am. Nat.* **127**, 1–8 (1986).
6. Clark, A. B. *Science* **201**, 163–165 (1978).
7. Gowaty, P. A. *Am. Nat.* **141**, 263–280 (1993).
8. Komdeur, J. *Behav. Ecol. Sociobiol.* **34**, 175–186 (1994).
9. Gowaty, P. A. *Curr. Ornithol.* **8**, 141–172 (1991).
10. Gowaty, P. A. & Droge, D. Sex ratio conflict and the evolution of sex-biased provisioning *Acta XX Congr. Int. Ornithol.* Vol. II, 932–945 (1991).

Figure 1 The Seychelles warbler — gem of a subject for studying adaptive sex ratios.

translocations between islands raises the work of Komdeur *et al.* from being extremely interesting to an instant classic. This is the first experimental demonstration that wild-living parents adaptively produce daughters and sons depending on the environmental situations in which they find themselves. When breeding pairs from poor-quality territories, who have previously produced sons, are moved to high-quality territories, they begin to produce daughters. In other words, when daughters are able to help, they are overproduced relative to sons; but when they can only hinder, they are under-produced.

An implication of the new study is that the effect of helper birds on the availability of food for their parents is a crucial selection pressure, whether this effect is to increase or to decrease local food supplies. As such, the availability of food or other resources is usually described as two different selection pressures — either local-resource enhancement or local-resource competition. But in fact, these may just be different sides of the same coin. The observations made by Komdeur *et al.*² further indicate that, when studying avian sex ratios within the context of natural selection³, it may be profitable to focus on the social situation in which individuals breed rather than on genes for all-son-producers or all-daughter-producers. Their study also indicates that empirical failures to show interesting sex-ratio variations in other socially monogamous, monomorphic bird species might be associated as much with the difficulty of assigning sex to young birds as anything else⁹.

This is a benchmark study that raises several important questions. First, how precise are the facultative adjustments of individuals? This question will only be answered with accurate data on individual variations in the mechanisms of resource enhancement and competition — the sort of data that Komdeur's living laboratory are likely to

Climatology

See-saw sea

James A. Carton

Variations in the Earth's climate over timescales of around a decade are particularly visible to us, and therefore well studied. One feature whose existence is still controversial is the Atlantic dipole — during some years, the northern tropical waters appear to warm anomalously, while the southern waters cool. In other years the pattern is reversed, the whole thing oscillating on a timescale of about 10 years. This pattern of warming and cooling leads to floods and droughts in Africa and the Americas. A clear demonstration of the physics of the oscillation has been missing from the debate over its existence, but on page 516,

Chang *et al.*¹ weigh in with a convincing mathematical model of the ocean and atmosphere. Perhaps rapid progress can now be made in explaining the decadal climate variability of the tropical Atlantic.

Tropical rainfall in the Atlantic region varies on timescales much longer than a year. Rainfall from the northern Nordeste region of Brazil is shown in Fig. 1a: during the mid-1970s and mid-1980s, rainfall was unusually heavy, whereas during the early 1980s there was a persistent drought. Explanations for this variability have focused on the atmosphere's response to observed variations in the sea surface temperature difference

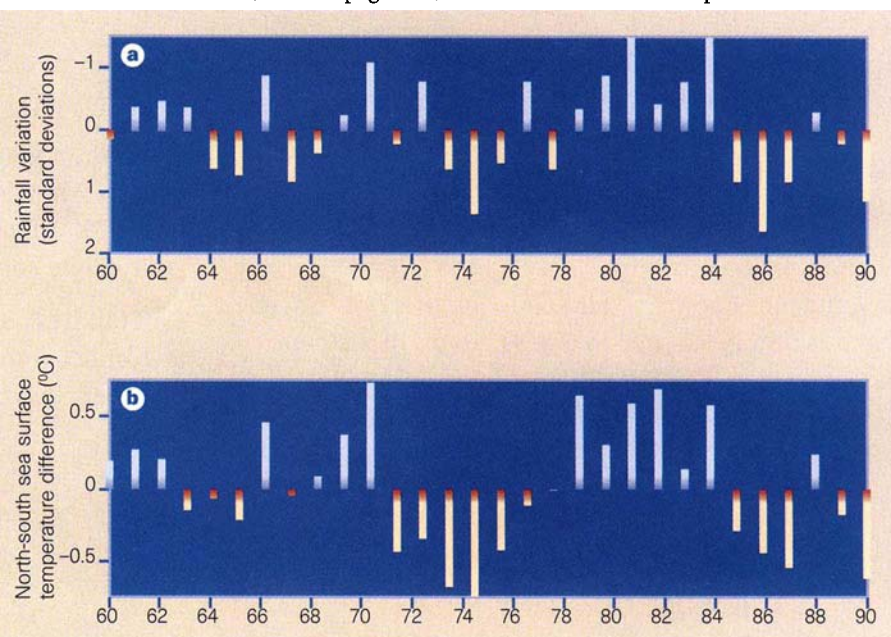


Figure 1 Skies shadow the sea: a, rainfall for the northern Nordeste region of Brazil (March–June) scaled by the yearly standard deviation²; b, the difference in sea surface temperature between the northern and southern tropics for March to May.

between Northern and Southern Hemispheres (the SST gradient).

Slight yearly changes in the SST gradient influence atmospheric circulation by their effect on surface air pressure^{2,3} and the direction of local time-averaged winds. North of the Equator, the northeasterly trade winds predominate, whereas south of the Equator the prevailing winds are from the southeast. A southward SST gradient (with Southern Hemisphere warmer than Northern) causes less southerly wind near the Equator. The temperature gradient will reduce air pressure in the Southern Hemisphere and increase it in the Northern Hemisphere. Owing to the effects of the Earth's rotation, clockwise circulations will be set up in both hemispheres, strengthening the northeasterly trade winds and weakening the southeasterly trade winds.

How does the ocean respond to these changing wind patterns? Decadal patterns of SST are mainly controlled by changes in evaporation, and that is due mainly to changes in surface wind speed⁴. So when the trade winds in the Northern Hemisphere strengthen and trade winds in the Southern Hemisphere weaken, SST will fall in the northern tropics and rise in the southern tropics, reinforcing the original gradient. That is what causes the instability.

This dominance of evaporation is in striking contrast to the El Niño phenomenon in the Pacific. El Niño depends mainly on fluctuations in the thickness of the warm upper layer of water within a few hundred kilometres of the Equator. These fluctuations are produced by equatorial Kelvin waves confined to the Equator by the effects of the Earth's rotation.

Chang *et al.*¹ present a coupled ocean-atmosphere model that captures much of the essential physics described above. The atmospheric behaviour is determined empirically, by statistically relating historical patterns of SST and corresponding patterns of wind. The ocean has a simplified vertical structure that retains its response to surface heat flux, and incorporates the effects of horizontal bulk transport. The authors show that horizontal transport may be the mechanism by which the dipole oscillates.

The strong South Equatorial Current carries water northwestwards out of the Southern Hemisphere, and the North Brazil Current and the North Equatorial Countercurrent then return this water to the interior of the Northern Hemisphere. So temperature anomalies will shift from one hemisphere to the other, eventually leading to a reversal of the pattern. Interestingly, an alternative mechanism relying on surface heat flux has been proposed by Zengxi Zhou⁵. In his model, weakened winds and increased SSTs near the Equator are associated with strengthened winds and decreased SSTs at higher latitudes. Gradually, the warm and cool anomalies shift towards the Equator,

replacing anomalies of opposite sign.

Might the success of this model lead to better long-range weather forecasting? The low-frequency nature of the Atlantic dipole and its relationship to broad patterns of SST explains the success of simple statistical forecasting schemes already in place. In these schemes, climate variables such as Brazilian rainfall are lag-correlated with a variety of environmental factors such as tropical SST. The statistical forecast is thus based on the empirically determined relationships. But statistical forecasting is limited because it doesn't depend on physical understanding. In work under submission, Chang *et al.* are already able to demonstrate that forecasts with this model improve on a persistence

forecast, in which the SST gradient is assumed to persist for long times⁶.

The success of this model will undoubtedly encourage research in the area of decadal climate variability. That should lead to rapid progress, not only on the Atlantic dipole, but on the potentially related problems of the mid-latitude ocean-atmosphere system such as the North Atlantic Oscillation. □

James A. Carton is in the Department of Meteorology, University of Maryland, College Park, Maryland 20742, USA.

1. Chang, P., Ji, L. & Li, H. *Nature* **385**, 516–518 (1997).
2. Hastenrath, S. & Greischar, L. *J. Climate* **6**, 734–758 (1993).
3. Hastenrath, S. & Greischar, L. *J. Geophys. Res.* **98**, 5093–5102 (1993).
4. Carton, J. A. & Huang, B. *J. Phys. Oceanogr.* **26**, 1165–1175 (1996).
5. Zhou, Z. thesis, Univ. Maryland (1996).
6. Chang, P., Ji, L. & Li, H. *Science* (submitted).

Evolutionary biology

How to keep a head in order

Per Erik Ahlberg

At first sight, the vertebrate head seems like an engineer's nightmare. Not only does it contain three separate skeletal systems — braincase, dermal bone and visceral arch — but it is animated by a veritable cat's cradle of muscles. And yet it works: through all the twists and turns of evolution, in animals as different as lampreys and chickens, the mechanical integrity of the head has been maintained. A paper by Georgy Köntges and Andrew Lumsden, published in *Development*¹, now reveals the mechanism that maintains this integrity, and promises to illuminate the origin of jaws — a key event in early vertebrate evolution.

It has long been known that the 'visceral arches' of the vertebrate head (that is, the jaws, middle-ear bones, tongue skeleton and gill arches) have a separate origin from the

braincase; the visceral arches are made from neural-crest cells, whereas the braincase is not. More recently it has become evident that the dermal bones of the head, which form under the skin surface (and include, for example, the bones of the face in humans) also derive from neural crest². Visceral arches and braincase are found already in hagfishes, the most primitive known vertebrates, but dermal bone is restricted to more advanced forms. Even so, dermal bone evolved at least 500 million years ago³, so all three are very ancient and fundamental components of the vertebrate body.

Neural-crest cells originate along the edges of the developing neural tube, but then stream down the sides of the head to reach the areas where they will form jaws, dermal bones and so on. An important part

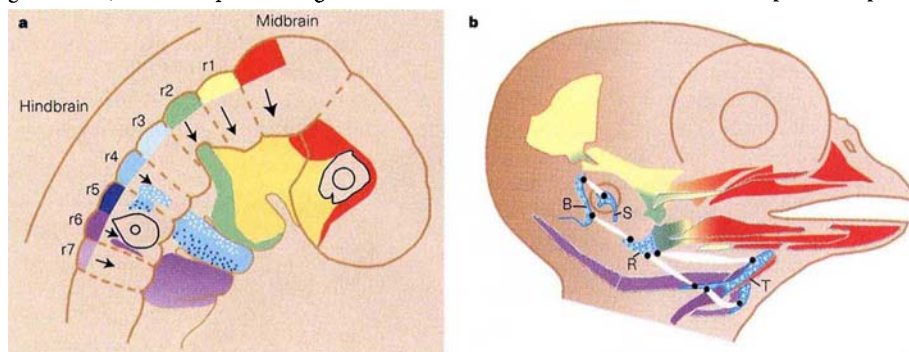


Figure 1 a, Early chick embryo (50–53 hours old) showing the seven rhombomeres of the hindbrain and the contributions they make to the developing visceral arches. Rhombomeres 3 and 5 kill most of their crest cells through apoptosis, but a few join the derivatives of rhombomere 4; these cells are indicated by light and dark spots. b, Advanced embryo (ten days old) showing bones of the upper and lower jaws, tongue skeleton and middle ear. These include both visceral and dermal elements. Muscles derived from rhombomere 4 are in white (other muscles have been omitted for clarity); note that the muscles always attach to bones of the same rhombomeric origin. The layout of the rhombomere components does not match the morphological boundaries of the bones. The boundaries of rhombomere 4 (blue) are very sharp, whereas rhombomeres 1–2 and 6–7 merge to some extent; this is probably determined by the apoptosis of rhombomeres 3 and 5. B, part of braincase; S, stapes; R, retroarticular process; T, tongue skeleton.

emanates from the hindbrain, where the neural crest is organized into seven neat segments called rhombomeres⁴⁻⁶ (Fig. 1a). It was natural to suspect that the rhombomeres might be involved in patterning the visceral arches, but until now it has been impossible to track the cells from any particular rhombomere to their ultimate destination.

Köntges and Lumsden solved the problem by replacing single chick rhombomeres with those of quails; because the two birds are closely related, the embryos develop normally, but the quail cells carry distinctive antigens and can readily be identified by antibody labelling even after they have dispersed into the head. So the contribution of any individual rhombomere to the developing head can be mapped, and an extraordinary picture emerges.

The authors show that rhombomeres generate coordinated 'packets' of muscle-and-skeleton, each packet consisting of the connective tissue of a muscle plus the two skeletal elements to which the muscle attaches. All the packets produced by one rhombomere are innervated by the same nerve. For example, rhombomere 4 (which is particularly easy to trace, as rhombomeres 3 and 5 on either side of it apoptotically kill their crest cells and contribute very little to the developing head) makes part of the tongue skeleton, the retroarticular process (which forms the rear end of the lower jaw in birds and reptiles, but not in mammals), and the muscles that link these two. It also makes the connective tissue of the jaw-opening muscle, which links the retroarticular process to the braincase, and — crucially — the small patch of braincase wall where this muscle attaches (Fig. 1b). All of these muscles are innervated by cranial nerve VII.

There are some major surprises here. First, nobody suspected that the braincase-wall attachments for the visceral-arch muscles were made of discrete neural-crest populations; without antibody labelling, they are indistinguishable from the surrounding mesodermal braincase wall. Second, the boundaries between rhombomere components in the visceral-arch elements do not correspond to the conventionally recognized segmentation of the branchial arches, which is based on visible morphology. The retroarticular process, for example, simply looks like part of the lower jaw and is never demarcated by a visible suture. But antibody labelling reveals a razor-sharp boundary between the process (made from rhombomere 4) and the rest of the jaw (made from an amalgam of rhombomeres 2 and 1, and midbrain neural crest).

In functional terms, the importance of this arrangement is obvious. The developmental programming of the rhombomeres ensures that the animal ends up with an orderly set of visceral arches, muscles which link them to the braincase, and proper

innervation of the muscles. This relationship of connections can be maintained even through dramatic changes in the shape and position of the various elements.

The evolutionary implications of this pattern are also exciting. Although direct evidence of the pattern is only available from birds at present (there is indirect supporting evidence from mice in which the homeobox gene *hoxa-2* has been knocked out), its extraordinary simplicity and power strongly suggest that it is a fundamental vertebrate characteristic. This means that it should be able to cast light on one of the most important but least understood events in vertebrate evolution: the origin of jaws.

Jaws are visceral arches, and thus broadly comparable to gill bars, but attempts to establish detailed serial homologies between jaws and gill arches have always run into trouble. Lampreys and hagfishes, the living jawless vertebrates, should show the primitive condition but have been little help until now; their gill arches are difficult to compare with those of jawed vertebrates, and they

have a very peculiar rasping tongue⁷.

Now, at last, we seem to have the key to a meaningful comparison. First, Köntges and Lumsden's work indicates that the precise innervation of lamprey and hagfish head muscles will provide important clues to their rhombomere derivation. Second, in the long run it should be possible to map the contributions of individual rhombomeres directly. That will show how the head structures of jawed and jawless vertebrates really relate to each other, and provide the best chance yet of understanding where jaws came from. □

Per Erik Ahlberg is in the Department of Palaeontology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK.

1. Köntges, G. & Lumsden, A. *Development* **122**, 3229–3242 (1996).
2. Couly, G. F., Coltey, P. M. & Le Douarin, N. *Development* **117**, 409–429 (1993).
3. Smith, M. P., Sansom, I. J. & Repetski, J. E. *Nature* **380**, 702–704 (1996).
4. Lumsden, A., Sprawson, N. & Graham, A. *Development* **113**, 1281–1291 (1991).
5. Lumsden, A. & Keynes, R. *Nature* **337**, 424–428 (1989).
6. Graham, A., Francis-West, P., Brickell, P. & Lumsden, A. *Nature* **372**, 684–686 (1994).
7. Janvier, P. *Early Vertebrates* (Oxford Science, Oxford, 1996).

Mineralogy

The Earth's mantle remodelled

Ross Angel

From petrological and geochemical data we know that the upper mantle of the Earth (about 100 to 660 km depth, see Fig. 1) is composed mainly of the common minerals olivine, pyroxene and garnet. But in what proportions? This long-standing issue seems close to resolution, as emerged at a meeting* held late last year.

The question of upper-mantle composition is by no means esoteric, for the answer would tell us a great deal about the convective regime of the mantle as a whole. If the upper mantle consists mostly of pyroxene and garnet (and about 40% olivine, the so-called 'piclogite' model), then there would be a strong chemical contrast between it and

the underlying lower mantle; that, in turn, would imply that the Earth convects as a double system with little or no mass exchange between the upper and lower mantle. But an olivine-rich upper mantle (about 60% olivine, in a 'pyrolite' model) would imply that there is no chemical difference between it and the lower mantle, and whole-mantle convection would be allowed.

The answer to the question can, in principle, be obtained by matching laboratory measurements of the characteristic seismic velocities of the various minerals to the seismic velocities obtained from geophysical observations. But over the past 15 years, as the experimental database of mineral prop-

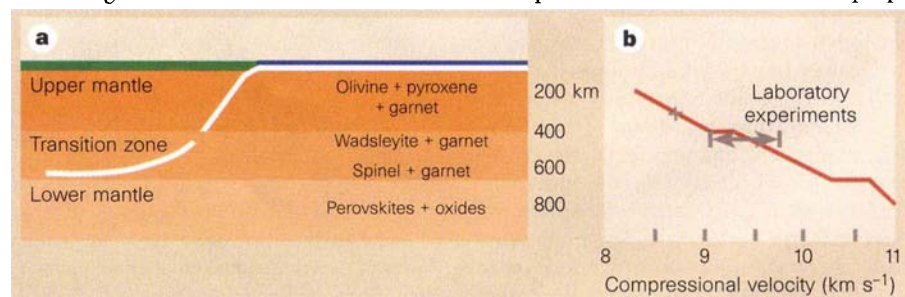


Figure 1 a, Overview of the structure of the upper 1,000 km of the Earth, showing the major minerals present in each region. Continental crust is on the left, and oceanic crust on the right; a subducting slab is depicted beneath the former. b, The variation of compressional wave velocity with depth derived from seismic observations (from ref. 3), together with the experimentally measured velocity of an olivine–pyroxene aggregate at mantle temperature and pressure (cross, from ref. 4). The velocity jump (arrow) for pure Mg_2SiO_4 at pressures corresponding to 410 km depth inferred from room temperature measurements (ref. 2, for example) is far greater than that observed in the Earth's mantle.

erties built up, it has become increasingly difficult to reconcile geochemically reasonable mineralogical models with the seismic observations (as discussed in News and Views just over a year ago¹). At the meeting, however, new data, arising from rapid developments in the pressures and temperatures at which experiments can be carried out in the laboratory, came under consideration. They point the way to a resolution of this problem in terms of a chemical stratification between the uppermost mantle (about 100–410 km depth) and the underlying transition zone (410–660 km).

At pressures and temperatures corresponding to a depth of about 410 km, olivine — a solid solution $(\text{Mg,Fe})_2\text{SiO}_4$ — transforms to a denser polymorph, wadsleyite (frequently and incorrectly termed 'β-phase' or 'modified spinel'). Laboratory measurements of the elasticity of both polymorphs of Mg_2SiO_4 (B. Li *et al.*, SUNY, Stony Brook; see also ref. 2) imply that velocity changes of some 13% accompany this transition. At the same depth within the Earth, geophysical models of the variation of seismic-wave velocities with depth show corresponding velocity changes of only 3–5% (for example, see ref. 3). From these data, the olivine content of the upper mantle has been inferred to be around 40%.

However, *in situ* measurements of olivine (R. C. Liebermann *et al.*, SUNY, Stony Brook) and olivine + pyroxene mixtures⁴ at high temperatures and pressures show that their seismic wave velocities exactly match those inferred for the upper mantle at depths of 200–400 km. Because garnet has much higher seismic velocities than olivine (for example, G. D. Gwanmesia *et al.*, Delaware State Univ.), and the high-density clinopyroxenes⁵ are inferred to be as fast as olivine, these data would imply a garnet-free mantle — a petrologically unacceptable model.

In consequence, it has been proposed instead that the 410-km discontinuity arises from a pyroxene-to-garnet transformation in a sodium-rich portion of the upper mantle which is a residue of an early magma ocean (T. Gasparik, SUNY, Stony Brook). Such a model would imply that little or no convection occurs in the upper mantle, a result that is difficult to reconcile with the existence of convection-driven plate tectonic activity on the Earth.

Resolution of these views comes from consideration of the latest data, and a chemically more complex model for the upper mantle and transition zone that involves the effects of both iron and water. Whereas $(\text{Mg,Fe})_2\text{SiO}_4$ olivine in isolation would transform to a wadsleyite phase of the same composition, in the presence of garnet it reacts to give rise to a wadsleyite enriched in Fe relative to olivine (Y. Fei and C. Bertka, Carnegie Inst. Washington). Such enrichment

in Fe reduces the seismic velocities in wadsleyite (S. Sinogeikin *et al.*, Univ. Illinois). Furthermore, wadsleyite accommodates far more hydrogen within its structure than does olivine, and its presence changes the crystallographic structure (J. Smyth, Univ. Colorado); the effects on the elastic properties are as yet unknown, although one can infer from crystal chemical considerations that there is an additional general softening of the structure and an accompanying decrease in seismic velocities. So the velocity jump at 410-km depth in the Earth from olivine to wadsleyite enriched in H and Fe will be less than that which would accompany the transition if the olivine and wadsleyite had the same composition. Therefore more olivine is now required than in the previous simple isochemical model to produce the same velocity jump.

Many of these data are too new to have been incorporated into geophysical models of the upper mantle. But calculations (A. R. Calderwood, Univ. British Columbia) suggest that a pyrolite upper mantle containing about 60% olivine, accompanied by Fe-enrichment in the higher density phases of the transition zone, gives a reasonably good match to modern global seismic data sets.

So debate over the mineralogical composition of the upper mantle now seems settled. But does the outcome mean, as previously assumed, that there is no chemical stratification between the transition zone and lower mantle, and thus that whole-mantle convection occurs? Perhaps not, because a direct result of an increased olivine content of the upper mantle is a greater release of latent heat from the olivine-to-wadsleyite transition in subducting slabs, which simulations suggest would encourage the physical detachment of these slabs at the 410-km discontinuity (D. Brunet, CNRS, Toulouse). The subsequent ponding of the detached cold slabs on the 660-km discontinuity that marks the top of the lower mantle would provide at least a partial physical barrier to convection; and it would increase the possibility of the mantle convecting as double system, with perhaps intermittent catastrophic overturns⁶ that mix the upper and lower mantle on a timescale of 100 million years or more. □

Ross Angel is in the Bayerisches Geoinstitut, Universität Bayreuth, D-95440 Bayreuth, Germany (e-mail: ross.angel@uni-bayreuth.de).

1. Jeanloz, R. *Nature* 378, 131–132 (1995).
2. Duffy, T. S., Zha, C.-s., Downs, R. T., Mao, H.-k. & Hemley, R. J. *Nature* 378, 170–173 (1995).
3. Kennett, B. L. N. & Engdahl, E. R. *Geophys. J. Int.* 105, 429–465 (1991).
4. Knoche, R., Webb, S. L. & Rubie, D. C. *EOS* 76, F563 (1995).
5. Angel, R. J., Chopelas, A. & Ross, N. L. *Nature* 358, 322–324 (1992).
6. Machetel, P. & Weber, P. *Nature* 350, 55–57 (1991).

*American Geophysical Union Fall Meeting, San Francisco, USA, 15–19 December 1996.

Daedalus

Visionary spectacles

Our eyes gradually lose their focusing ability with age. Reading glasses, bifocals and graded-focus spectacles are a poor substitute. Daedalus now has a better one.

His first idea was a hollow lens whose interior could be filled with a liquid. Replace that liquid by one with a different refractive index, and the lens would change its focal length. But a multi-focus lens with a range of different liquids would pose fearsome plumbing problems. A refractive gas, pumped into the lens cavity under variable pressure, seems simpler. Sadly, most gases change their refractive index only slightly with pressure.

A supercritical fluid, however, could be pressured from an index of almost unity right up towards that of glass itself. But the pressures needed could be rather high. Supercritical xenon, for example, exerts about 60 atmospheres at its index peak; adjusted for reading, xenon-focused glasses would be a potential bomb. But Daedalus recalls how readily gases condense on absorbents such as silica gel, even at quite low pressures. And the ultimate silica gel is silica aerogel, that amazing open silica framework not much denser than air, and indeed known as 'frozen smoke'. With its vast internal surface, it must absorb gases avidly. Sadly, in its pure state it is not quite optically clear. But Daedalus argues that a little adsorbed gas will make it so. Refractive matching will reduce the scattering from its internal microstructure.

So Daedalus's new variable-focus 'Aerospecs' will have hollow lenses filled with silica aerogel, and connected to a supply of an easily condensed gas such as butane, ammonia or a suitable halocarbon. For distance viewing, the lenses contain just enough gas to keep the aerogel transparent. For closer viewing, a slight rise in pressure condenses more gas into the aerogel, and the lenses shorten their focus.

Aerospecs will be quite automatic. An auto-focus system, of the type used in many cameras, will be mounted on their frame. It will stare forward into their field of view, and will drive the reversible gas-pump that adjusts their lenses. All this machinery will be hard to pack into even the most overbearing frames; but fortunately, high accuracy is not needed. Fine focusing will be done automatically by the wearer's own eyes. He will feel that his youthful accommodation has been magically restored. One problem may be that when gas is pumped rapidly out of the lenses, desorptive cooling will mist them up. But spectacle wearers are already resigned to troubles of that sort.

David Jones

Alex Todd (1907–97)

Organic chemist, and mover-and-shaker in British science policy

Alexander Robertus Todd, Scotsman and Nobel prizewinner, who died on 10 January in Cambridge, England, at the age of 89, was a big man in every sense. Height made him tower above others. Quick wits and impulsive generosity made him lifelong friends. Impulsive intolerance made him enemies as well. As a young man, he was a prodigy; later, he was simply indefatigable.

The twentieth century has been well served by its synthetic organic chemists, from Emil Fischer onwards. Todd, the son of a line-manager on the Glasgow tramway system, was a precocious undergraduate at Glasgow and afterwards spent two years at Frankfurt, earning a PhD (in German) for his work on bile acids. But he learned his craft from Robert Robinson, Oxford's legendary green-fingers, with whom he worked in the early 1930s before shuttling as a postdoctoral student between Edinburgh and the Lister Institute in London, ticking off structures and syntheses as he went: vitamin B₁, α - and β -tocopherol and, to his later delight, the active principle of cannabis (cannabinol).

He became head of the chemistry department at the University of Manchester in 1938, at the age of 31, where he embarked on the study of the purine and pyrimidine bases in living things. In 1944, he became head of chemistry at Cambridge: there his group produced the structures of vitamin B₁₂ and contributed to that of penicillin, and also synthesized ATP.

Todd's Nobel prize was awarded in 1957 for his work on purines and pyrimidines, which had provided essential information for the model-building that eventually enabled Watson and Crick to produce a plausible structure for DNA. Todd and his collaborators showed that the deoxyribose in nucleotides is cyclized, that the link between the sugar and the base is a β -linkage and that, in polymers, the phosphate groups are invariably linked to the 5' and 3' carbons of the sugar molecule (giving a single strand of DNA its crucial directionality). It is odd that he himself never took up model-building.

By the late 1950s, Todd had also begun to sense the pleasures of public life. He cut his teeth on the struggle to build a new chemistry laboratory in Cambridge (completed in 1958), when he seems to have learned that a person with a clear idea of

what he wants can often get it by requesting it in a variety of tones of voice ranging from the clear and reasonable to the determined and even threatening. (Height helped again.)

Todd was proud of being a Scot and of his achievements as a chemist, of course, but there were two other feathers in his cap that kept cropping up in conversation. One was his chairmanship of the Advisory Council on Scientific Policy between 1952 and 1964, whose job was to advise the British government on the spending of research funds, but which stood out (and still does) among such committees in publishing an often critical annual report. How did that come about? "I asked a civil servant to arrange it", he once said; one sympathized with the poor man concerned.

He also played a central part, as a trustee of the Nuffield Foundation, in launching the programme of curriculum development for schools called the Nuffield Science Teaching Project (of which I was the coordinator from 1964 to 1966). On one occasion, it fell to me to ask for an extra £2 million for several new projects and to Todd to speak first. "Can any trustee say the foundation's ever done a better thing?" was almost all he said. We got the funds.

In 1963, Todd — by now a life peer, and as such Lord Todd — was elected Master of Christ's College, Cambridge. Although still active in the laboratory, he seemed to have more time to spare for making waves. He was out of sympathy with many things: the Robbins report of 1963 recommending an expansion of British universities, the Trend report on the organization of British science, and even the election of the Labour government in 1964. "They think I'm to the right of Ghengis Khan", he used to say with relish.

Despite that insouciance, Todd was easily hurt when his views on public policy did not carry all before them. His presidential address to the British Association in 1970 (delivered in Glasgow Cathedral) was politically incorrect in many ways, advocating among other things the preferential direction of research funds to elite universities, much as the funding councils have now decided, and was derided by several newspapers. He seemed genuinely puzzled, over a consolatory whisky, that he had stirred up a

wasps' nest. Making public policy was one thing, controversy quite another.

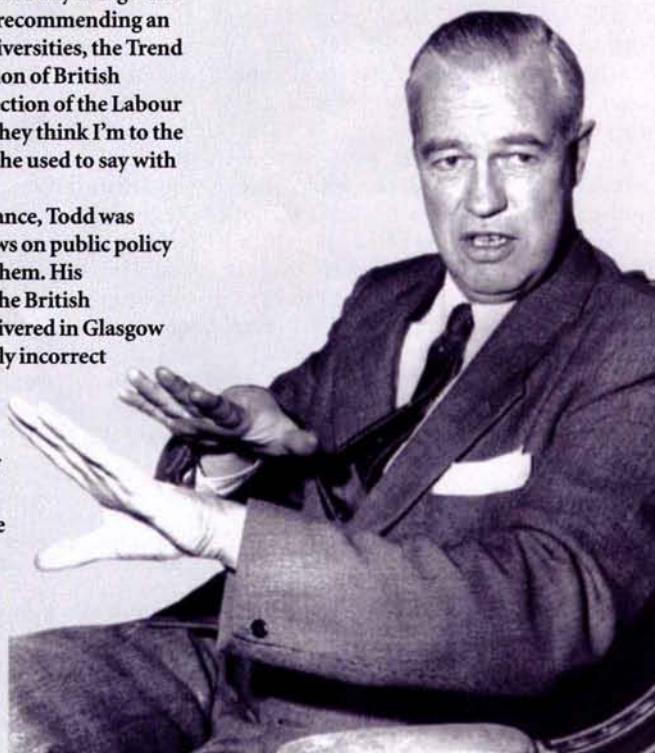
By then, the travel bug had bitten Todd. He enormously enjoyed travelling, more for the people he met than for the sights or the grand occasions. His photograph album must be full of groups of people at airports and dining halls, from Tokyo to Tashkent, in which the centrepiece is this two-metre Scot.

Travel caused an interesting fuss in 1975, when Todd was about to be elected president of the Royal Society. I was then director of the Nuffield Foundation (after a first spell at *Nature*). One day in October, he dropped by to explain that there would be so much travelling in the new post that he would have to resign as chairman of the trustees, so could I canvass the others to nominate a successor. We settled on Lord Trend (whose report, when he was secretary to the British cabinet, had ironically put an end to Todd's advisory council a dozen years earlier), who was promptly elected deputy.

And then Todd decided that he could fit in the travelling after all, and stayed for a further five years. (A little later, he was off to Hong Kong every other month to run a Nuffield look-alike called the Croucher Foundation, which was financed by an old friend and shipping magnate.) I said that it was for him to explain to Trend; but the explanation, I would guess, was more a declaration. Todd was like that. It is not so much that he was sure of himself as that he concealed self-doubt even from himself.

John Maddox

John Maddox is Emeritus Editor of Nature.



Politicians' uniquely simple personalities

The complexity of human personality has been reduced to five dimensions, based on factor analyses of judgements of personality traits¹. Many researchers agree that a five-factor model of personality captures the essential features of all traits that are used to describe personality: energy/extroversion; agreeableness/friendliness; conscientiousness; emotional stability against neuroticism; and intellect/openness to experience²⁻⁴. But we show here that this common, standard set of five factors does not hold for judgements of famous political figures.

We found that, when people judge the personality traits of politicians, they use only two or three factors. Personality factors that are normally independent — such as energy and openness — were highly correlated in a more simplified view of personality.

Political candidates gain intense media exposure over an extended period of self-promotion designed to portray them as trustworthy experts with many admirable personality traits⁵. Such public exposure is intended to lead to clearly articulated perceptions rather than stereotypical evaluations by the electorate⁶.

The nature of campaign information is unique as a basis for forming impressions of personality, as it is packaged by supporters and opponents as pros and cons (favourable or condemning) designed to simplify the ultimately dichotomous decision of how to vote. The selective mental processing and filtering by the electorate of the mass of discrepant input about political candidates must in the end justify each per-

son's one vote: be it for or against. Therefore, we predicted that personality judgements about political candidates would likewise be constricted to involve a limited number of factors rather than the usual five.

We first studied the personality judgements of a sample of 2,088 Italian adults, of diverse ages, education and political views. Leading party politicians were evaluated by 1,257 respondents, and another 831 evaluated their own personalities and those of several celebrities. Judgements were made from a list of 25 adjectives that are markers of the five-factor model. Each adjective (for example, enterprising, reliable, truthful) was rated on how characteristic it was of each target on a seven-point scale, and those ratings were factor-analysed⁷. The analysis reduces the scores to a minimal number of correlated groups of traits within factors that are independent of each other.

Ratings were made of two Italian political candidates (Silvio Berlusconi and Roman Prodi), an international celebrity (skiing hero Alberto Tomba) and a famous Italian television personality (Pippo Baudo).

Table 1 reveals three clear results: (1) respondents' personality portraits of themselves require the five-factor solution, as found in earlier research; (2) personality judgements of national celebrities also require five factors; but (3) personality judgements of political candidates are drastically reduced to only two factors, despite many significant differences between their personalities.

The first of the two stable⁸ personality factors for politicians has been named energy/innovation (which is a blend of energy and openness), and the second factor is honesty/trustworthiness (a blend of agreeableness, conscientiousness and stability).

These findings can be applied more generally, as shown by our replication study with 195 US college students. These students rated their own personalities after having rated Democratic President Bill Clinton and Republican presidential candidate Bob Dole, along with basketball star 'Magic' Johnson. The same 25 five-factor model marker adjectives were used as in our Italian study.

Table 1 shows that this different sample replicates the basic factor patterns found in the larger Italian sample: self-ratings and the ratings of the popular basketball player (among basketball fans) use all five factors, but judgements of the politicians are restricted to only three factors (among potential voters).

Finally, the percentage of total variance explained by each factor solution (2, 3 or 5) for each target personality, for both samples, is a high, nearly identical, average of 60 per cent.

We conclude that, by adopting a simplifying method of judging political candidates' personalities, voters use a cognitively efficient strategy for coding the mass of complex data, thus combating informational overload⁹. Doing so helps them to decide how to cast their vote.

Gian Vittorio Caprara*

Claudio Barbaranelli

University of Rome,

"La Sapienza",

Via dei Marsi 78,

Rome 1-00185, Italy

Philip G. Zimbardo

Psychology Department,

Stanford University,

Stanford,

California 94305, USA

e-mail: zim@psych.stanford.edu

Table 1 Factor composition for target personalities

	Factors					Variance explained (%)
	1	2	3	4	5	
Italian sample						
Self (<i>n</i> = 827)	E	O	A	C	S	56
Athlete (<i>n</i> = 829)	A	E	C	O	S	57
TV star (<i>n</i> = 830)	C	E	S	O	A	60
Politicians						
Berlusconi (<i>n</i> = 1,257)	A + C + S	E + O	–	–	–	61
Prodi (<i>n</i> = 643)	A + C + S	E + O	–	–	–	64
US sample						
Self (<i>n</i> = 195)	A	C	E	O	S	57
Athlete (<i>n</i> = 81)	S	C	O	A	E	61
Politicians						
Clinton (<i>n</i> = 127)	E + O	C + A	?	–	–	57
Dole (<i>n</i> = 127)	C + A	?	E + O	–	–	62

Factors 1-5 are arranged in order of the amount of variance in ratings, with Factor 1 explaining the most variance and Factor 5 the least. E, energy; O, openness; A, agreeableness; C, conscientiousness; S, emotional stability; and '?', an uninterpretable factor. Adjectives used to describe the politicians' two factors, normally attributed to the factors shown in parentheses, are — **energy/innovation**: enterprising (E), active (E), self-assured (E), energetic (E), cheerful (E), innovative (O), creative (O), inventive (O), smart (O), modern (O), efficient (C), optimistic (S), confident (S), cordial (A); and **honesty/trustworthiness**: sincere (A), truthful (A), loyal (A), responsible (C), reliable (C), precise (C), persistent (C), poised (S), peaceful (S), stable (S), generous (A).

1. Costa, P. T. & McCrae, R. R. *The NEO Personality Inventory Manual* (PAR, Odessa, 1985).

2. Briggs, S. J. *Personality* **60**, 254-293 (1992).

3. Caprara, G. V., Barbaranelli, C., Borgogni, L. & Perugini, M. *Personality Individ. Diff.* **15**, 281-288 (1993).

4. Goldberg, L. R. *Am. Psychol.* **48**, 26-34 (1993).

5. Simonton, D. K. *Why Presidents Succeed: A Political Psychology of Leadership* (Yale Univ. Press, New Haven, 1987).

6. Pierce, P. *Political Psych.* **14**, 21-35 (1993).

7. Cattell, R. B. & Vogelmann, S. *Multivariate Behav. Res.* **12**, 289-325 (1977).

8. Tucker, L. R. *A Method for Synthesis of Factor Analysis Studies* (Dept of the Army, Washington DC, 1951).

9. Fiske, S. & Taylor, S. *Social Cognition* (McGraw Hill, New York, 1991).

*More detailed methods and additional results are available from G. V. C. at caprara@axrma.uniroma1.it

Synergy between synthetic oestrogens?

Arnold *et al.*¹ recently reported that a range of chemicals could act synergistically to bind and activate the human oestrogen receptor (hER) *in vitro*. Most markedly, a 1:1 mixture of the insecticides dieldrin and endosulfan produced a 1,000-fold higher activity than when present alone. Arnold *et al.*¹ interpreted this as support for a role of environmental oestrogens in increasing the incidence of human breast and testicular cancers, and decreasing male sperm counts and sperm quality. The chemical selection criteria and the test data have subsequently been discussed and challenged²⁻⁵. We have evaluated the oestrogenic effects of dieldrin and endosulfan directly using two standard assays. Our data fail to confirm the conclusions of Arnold *et al.*

We used two assays that are sensitive to a wide range of oestrogens. In a yeast hER

transactivation assay⁶, the hER sequence is incorporated into the yeast genome. Activated oestrogen receptors are detected colorimetrically by a *lacZ* reporter gene. In a uterotrophic assay, sexually immature rats are treated with test chemicals on three successive days, and uterus mass is measured on the following day. Oestrogenic chemicals cause premature growth of the uterus⁷. This latter assay is generally regarded as the most reliable test for oestrogenic activity.

The top dose levels used in the *in vitro* assays were determined by the solubility of the agents in the assay medium and (in the uterotrophic assay) by clinical observations establishing systemic absorption of the test agents and achievement of a maximum tolerated dose. The low dose level of 5 mg kg⁻¹ evaluated in the uterotrophic assays was to take account of the possibility that synergism might only be evident with low levels of agonist bound to the receptor. The lowest of the dose levels used in the *in vitro* assays automatically took account of that possibility.

Dieldrin, endosulfan, and a 1:1 mixture of each chemical, were inactive as oestro-

gens in both assays. These results also established an absence of synergism (Fig. 1). A suppression of gain in body mass caused by the chemicals accounted for the reduced uterus mass seen in some of the uterotrophic assays. These negative results do not support the findings of Arnold *et al.*¹, although it is possible that the lack of oestrogenic activity of dieldrin and endosulfan may have compromised any observation of synergism between them.

Consequently, we evaluated the potential for synergism between dieldrin and methoxychlor, an established oestrogen⁷ (Fig. 1). For the dieldrin/methoxychlor mixture experiments a dose level of methoxychlor was selected that gave a sub-maximal oestrogenic response in each of the assays. No synergism was observed in either of the assays (Fig. 1).

Our results do not support the assertion that synergism between oestrogens is likely to present a major human or wildlife health concern.

J. Ashby, P. A. Lefevre, J. Odum

Zeneca Central Toxicology Laboratory,
Alderley Park, Cheshire SK10 4TJ, UK

C. A. Harris, E. J. Routledge, J. P. Sumpter

Department of Biology and Biochemistry,
Brunel University, Uxbridge,
Middlesex UB8 3PH, UK

1. Arnold, S. F. *et al.* *Science* **272**, 1489–1492 (1996).
2. Joffe, M. *Science* **274**, 1285 (1996).
3. McLachlan, J. *Science* **274**, 1285–1286 (1996).
4. Ramamoorthy, K. *et al.* *Science* **275**, 405 (1997).
5. McLachlan, J. A. *et al.* *Science* **275**, 405–406 (1997).
6. Routledge, E. J. & Sumpter, J. P. *Environ. Toxicol. Chem.* **15**, 241–248 (1996).
7. Shelby, M. D. *et al.* *Environ. Health Perspect.* **104**, 1296–1301 (1997).

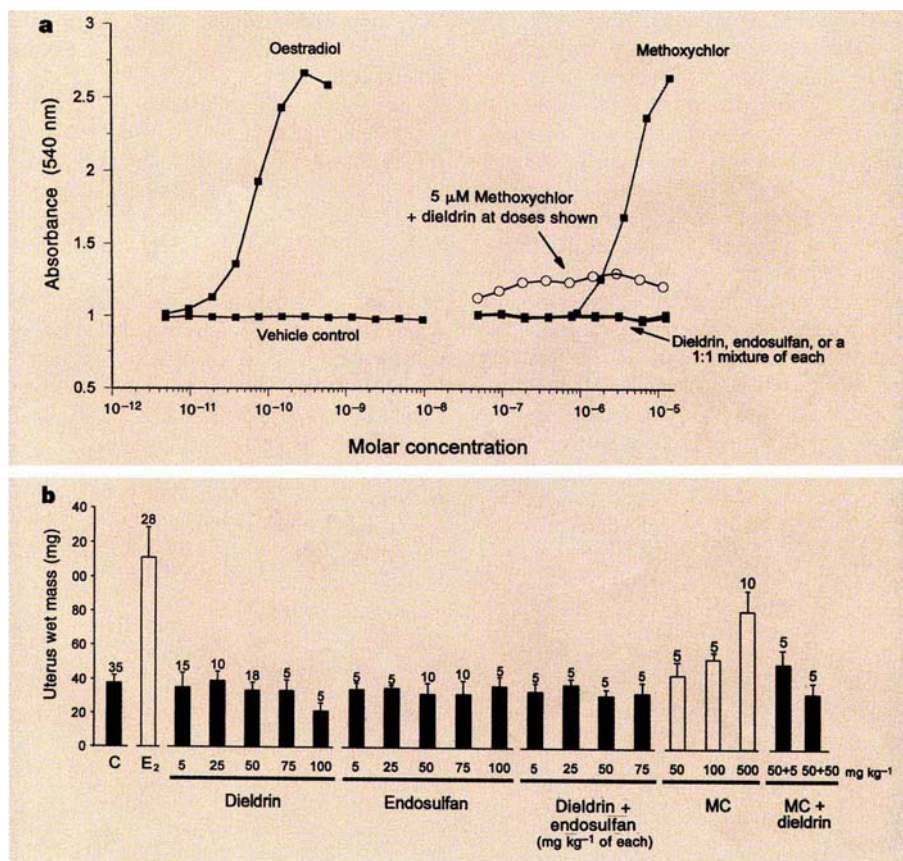


Figure 1 a, Results of the yeast oestrogenicity assay⁶. Several different commercial samples of the three test chemicals were evaluated, with identical results. Representative data are shown for a sample of dieldrin (Chem Service, West Chester, PA; 98%), a sample of endosulfan (Chem Service; 98%, 60% α -isomer, 38% β -isomer), and a sample of methoxychlor (Sigma; 98%). Test agents were dissolved in ethanol and added to 96 well plates at the concentrations shown. Plates were incubated with the yeast for 4 days at 32 °C (ref. 6). b, Immature Alderley Park (Wistar-derived) rat uterus mass after three daily subcutaneous injections (5 ml kg⁻¹) of arachis oil (control; C), oestradiol (E₂, 50 µg) or the agents shown at the dose levels indicated. Mean results ($\pm$ s. d.) are shown with animal group sizes above the bars. The results of several experiments are shown. MC, methoxychlor. The positive control data are shown as open bars. Samples were as in a.

Ant soldiers are not modified queens

On the basis of certain similarities between ant soldiers and the queen (gyne) of the same species, Baroni Urbani and Passera¹ propose that soldiers evolved as a caste independent of workers, and that they originated directly from the queen. Their argument is undermined by a large body of comparative evidence supporting the conventional view that the existence of discrete minor worker and major worker (soldier) morphs simply represents the end point of an evolutionary sequence that begins with the presence of weakly polymorphic workers and passes through an intermediate phase with an expanded, and progressively bimodal, worker-size frequency distribution²⁻⁵. Morphological divergence between the worker castes is further accentuated by allometric growth³.

For each of the three ant genera cited by Baroni Urbani and Passera (*Camponotus*, *Pheidole* and *Zacryptocerus*), in which some

species have a discrete soldier caste, there are related taxa in the same genus showing continuous polymorphism, that is, species with minor workers, major workers and at least some (often many) individuals intermediate in morphology between majors and minors⁵⁻⁷. In contrast, there are no species in which soldier-queen intermediates occur with any regularity. Continuous worker polymorphism appears to be the predominant condition in *Camponotus*, including the three species (*C. aethiops*, *C. ligniperdus*, *C. vagus*)⁸ whose egg-laying soldiers are discussed by Baroni Urbani and Passera¹. (Incidentally, their use of the term "soldier" for these three species contradicts their contention that the largest individuals of continuously polymorphic species are not true soldiers.)

Those relatively few *Camponotus* species that have a discrete soldier caste are a small subset of this large (>900 species) genus^{9,10}. Other genera in the tribe Camponotini have monomorphic or weakly polymorphic workers^{4,10}. A comparable situation exists in *Zacryptocerus*². Despite uncertainty about the exact morphological complexion of the ancestral, monomorphic worker caste^{11,12}, the nested occurrence of character states clearly supports the prevailing model of worker caste evolution from monomorphism through continuous polymorphism to discrete polymorphism.

In those species of ants with a discrete soldier caste it is possible to recognize some features shared with the queen, but in other traits (degree of wing development, number of thoracic sclerites, defensive behaviour) the soldier is decidedly more like the minor worker caste. Because of size-related allometric effects, one might expect that soldiers approach the queen morphotype more closely than do the minor workers; but that is not equivalent to the claim that soldiers are more similar to queens than they are to minor workers. Baroni Urbani and Passera confuse these two issues.

The extent of similarity between soldiers, queens and minor workers will depend on which characters and taxa are chosen. Some patterns of morphometric variation in the ant genus *Pheidole*, for example, reveal a closer resemblance between soldiers and minor workers than between soldiers and queens (Fig. 1). This is consistent with the data on caste determination in *Pheidole* indicating that queen-worker differentiation occurs early in development, whereas the fate of a worker larva — whether it becomes a minor worker or a soldier — is determined later, in the final larval instar^{13,14}.

Mere resemblance between soldiers and queens in some or even many traits does not demonstrate a queen-based origin of the soldiers. The case of ergatoid (wingless, worker-like) males is an instructive one. In several ant genera, including *Hypoconer*, *Cardiocondyla* and *Formicoxenus*, the males have evolved a strikingly worker-like morphology, in association with a syndrome of intranidal mating and strong male-male competition^{5,15-17}. In some instances both typical winged males and ergatoid males occur in the same species^{15,18}. In most respects the ergatoid males are morphologically much more similar to the worker caste than to the winged form. But it would be specious to argue that such males originated from workers. A much more compelling hypothesis is that males evolved a flexible developmental programme, such that one of two possible ontogenetic sequences involves de-repression of genes for worker morphology (which are, of course, present in all male ants but usually not expressed).

Similarly, the evolution of a polymorphic worker caste in ants may lead to the larger workers expressing some queen-like traits where this is selectively advantageous, at either the colony level (for example, plug-shaped heads) or perhaps the individual level (for example, increased fertility). The

efficacy of selection on worker phenotypes and the extent to which different (sub)castes of a species can evolve independently remain intriguing problems^{11,19}, but comparative morphological data, as well as the known facts about caste determination, point unequivocally to the origin of soldiers within the worker developmental pathway.

Philip S. Ward

Department of Entomology,
University of California,
Davis, California 95616, USA
e-mail: psward@ucdavis.edu

1. Baroni Urbani, C. & Passera, L. *Nature* **383**, 223 (1996).
2. Emery, C. *Biol. Centralbl.* **14**, 53–59 (1894).
3. Wilson, E. O. *Q. Rev. Biol.* **28**, 136–156 (1953).
4. Oster, G. F. & Wilson, E. O. *Caste and Ecology in the Social Insects* (Princeton Univ. Press, 1978).
5. Hölldobler, B. & Wilson, E. O. *The Ants* (Harvard Univ. Press, Cambridge, MA, 1990).
6. Creighton, W. S. & Gregg, R. E. *Univ. Colorado Stud. Ser. Biol.* **3**, 1–46 (1955).
7. Kempf, W. W. *Rev. Entomol.* **22**, 1–244 (1951).
8. Bernard, F. *Les Fourmis d'Europe Occidentale et Septentrionale* (Masson, Paris, 1967).
9. Emery, C. *Genera Insectorum* **183**, 1–302 (1925).
10. Bolton, B. *A New General Catalogue of the Ants of the World* (Harvard Univ. Press, Cambridge, MA, 1995).
11. Wilson, E. O. *Behav. Ecol. Sociobiol.* **7**, 143–156 (1980).
12. Wheeler, D. E. *Psyche (Camb.)* **91**, 171–192 (1985).
13. Passera, L. in *Caste Differentiation in Social Insects* (eds Watson, J. A. L., Okot-Kotter, B. M. & Noirot, C.) 331–346 (Pergamon, Oxford, 1985).
14. Wheeler, D. E. *Am. Nat.* **128**, 13–34 (1986).
15. Le Masne, G. *Insectes Soc.* **3**, 239–259 (1956).
16. Stuart, R. J., Francoeur, A. & Loiseleur, R. *Naturwissenschaften* **74**, 548–549 (1987).
17. Yamauchi, K., Kimura, Y., Corbara, B., Kinomura, K. & Tsuji, K. *Insectes Soc.* **43**, 119–130 (1996).
18. Bolton, B. *Bull. Br. Mus. Nat. Hist. Entomol.* **45**, 307–370 (1982).
19. Wheeler, D. E. *Am. Nat.* **138**, 1218–1238 (1991).

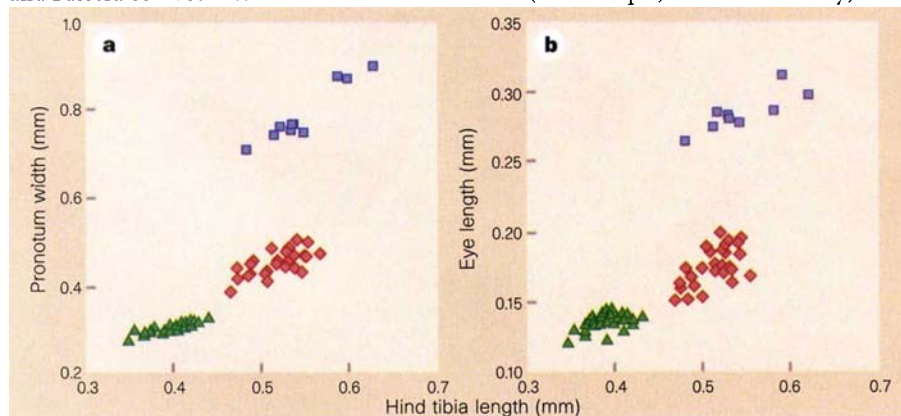


Figure 1 Bivariate plots of morphometric characters in a random selection of queens (blue squares, $n = 10$), minor workers (green triangles, $n = 53$) and soldiers (red diamonds, $n = 32$) from five colonies of *P. californica* from the vicinity of Davis, California. The soldiers are closer to the queens in body size (as measured by the length of the hind tibia), but the relationships of (a) soldier pronotum width and (b) eye length to hind tibial length are much more similar to those of the minor workers.

HIV-1 tropism and co-receptor use

Chemoattractant cytokines (chemokines) act during inflammation by binding to and signalling through 7-transmembrane cell-surface receptors. There has been much excitement over the discovery that certain chemokine receptors, in cooperation with CD4 molecules on the T-cell surface, also serve as entry portals for HIV¹. Various isolates of HIV-1 use different chemokine co-receptors. This helps to explain why some HIV isolates preferentially infect one or other cell type, a property known as HIV tropism. Although almost all HIV isolates can infect T-helper lymphocytes that have been activated to proliferate in primary culture, only certain isolates can also infect macrophages, whereas others infect immortalized T-cell lines. We show here that it is oversimplistic to assume that chemokine co-receptor use accurately reflects the cellular tropism of the virus.

Recent evidence shows that the chemokine receptors CCR5 and CXCR4 are the most important co-receptor molecules

required, in addition to CD4, for entry of macrophage-tropic and T-cell-tropic HIV-1 isolates, respectively²⁻⁴. HIV-1 isolates that show a tropism both for T-cell lines and macrophages can use either of these receptors⁵⁻⁷, whereas another chemokine receptor, CCR3, can accommodate a small subset of macrophage-tropic isolates⁸. The homozygous deletion in the CCR5 gene abolishes the infectivity of viruses that use only CCR5 as a co-receptor in T-lymphocytes and macrophages⁹.

The tropisms of HIV-1 act independently of genotype or subtype clade^{5,8}, but are determined by the variable domains V1/V2 and V3 of the outer-envelope glycoprotein (gp120) of HIV. Neutralizing antibodies specific to the V3 loop of primary HIV-1 isolates inhibit the interaction of gp120 with the CCR5 co-receptor^{10,11}.

Despite the excellent data on the role of chemokine co-receptors in HIV entry and cell tropism, we have found HIV-1 strains that clearly use CCR5 and/or CCR3 in transfected cell lines yet cannot infect macrophages, even when all non-envelope genes are of the permissive genotype. One needs to beware of the emerging assumption that chemokine co-receptor use accurately reflects the cellular tropism of the virus, and that all 'non-syncytium-inducing' strains (those that do not cause formation of multinucleated cells) are macrophage-tropic.

A 'syncytium-inducing' isolate of HIV-1 (Gun-1 wild type) shows dual tropism, in that it can efficiently infect both T-cell lines and primary macrophages¹². Selection of this isolate to infect CD4⁺ brain cells resulted in a variant (Gun-1 V) that maintained its infectivity for T-cell lines but not for macrophages. This change in tropism is determined by a single amino-acid change in the envelope, specifically in the V3 loop¹³.

Because the V3 loop is a target for neutralizing antibodies as well as being a determinant of cellular tropism, we produced a panel of potent neutralizing monoclonal antibodies to a V3 loop peptide of Gun-1 V and used it to select for four neutralization-escape mutants. We tested the resulting isolates (V.escape 1-4) for their tropism.

We have shown previously that substitution of amino acids in the V3 loop in all of the neutralization-escape mutants was concomitant with altered tropism¹². We have now analysed all six isolates for co-receptor use by testing their ability to infect a CD4⁺ feline kidney-cell line (CCC) transiently expressing the human chemokine co-receptors CXCR4, CCR5 or CCR3 (Table 1). As expected, the dual-tropic Gun-1 wild-type isolate can use either CXCR4 or CCR5, whereas Gun-1 V can only use CXCR4 efficiently. This result is consistent with the current model of co-receptor use. However, when the V3 loop neutralization-escape mutants are tested, this model fails. Two of the escape mutants show efficient use of the macrophage co-receptor CCR5 expressed on CCC cells, but do not efficiently infect macrophages.

We also obtained similar results (Table 1) with primary molecular clones, in this case a non-syncytium-inducing phenotype, derived from blood (E4), lymph node (LN15), spleen (S8), brain (B8) and lung (L5) tissue of one patient with AIDS¹⁴. The clones were isolated by long-range polymerase chain reaction without any selection in cell culture. Tropism studies revealed that these virus clones are only infectious for peripheral blood mononuclear T-cells, and all replicated with comparable kinetics. They do not infect macrophages.

Co-receptor studies using a single-round infectivity assay showed that the

envelopes of these primary molecular clones are able to use CCR5 and CCR3 transiently transfected into CD4⁺ CCC cells. Quantitative estimates of the efficiency of CCR5 and CCR3 use (on transfected CCC cells) by the viruses described here are equivalent to those of Gun-1 wild type which efficiently infects macrophages. Our observations therefore cannot be explained by differential CCR5 expression.

So five molecular clones (non-syncytium-inducing) and two biological clones (syncytium-inducing) are all able to use the 'macrophage co-receptors' CCR3 and/or CCR5, yet they are unable to use these molecules when they are expressed by macrophages. It is also apparent that macrophages express CXCR4 (ref. 15), yet are not infected by T-cell-line-adapted strains of HIV-1. We conclude that the ability of a virus to use a co-receptor molecule does not necessarily confer on that virus the ability to use that molecule on all cell types. Post-translational modifications of co-receptors on macrophages may affect their availability for infection by some HIV-1 strains but not others.

We agree with Zhang *et al.*⁵ that the evolution of HIV-1 from non-syncytium-inducing to syncytium-inducing strains, by gaining the ability to infect cells via CXCR4, will not necessarily result in the loss of CCR5 use⁷. We have shown that the selective pressure on HIV-1 to immune escape from neutralization can result in gaining the use of co-receptors. Conversely, selection to use these new co-receptors, as the virus colonizes various cell types *in vivo* during the progression to AIDS, may result in escape from neutralizing antibodies.

Matthias T. Dittmar, Áine McKnight, Graham Simmons, Paul R. Clapham Robin A. Weiss

Chester Beatty Laboratories,
Institute of Cancer Research,
237 Fulham Road,
London SW3 6JB, UK

Peter Simmons

Department of Medical Microbiology,
University of Edinburgh Medical School,
Teitot Place, Edinburgh EH8 9AG, UK

Table 1 HIV-1 co-receptor use and cellular tropism

HIV-1 strain	Syncytial phenotype	Tropism*			Co-receptor use†		
		T-cell lines	PBMC	Macrophage	CCR3	CCR5	CXCR4
HXB2	SI	+	+	-	-	-	+
Gun-1 wild type	SI	+	+	+	-	+	+
Gun-1 V	SI	+	+	-	-	-/+	+
V.escape 1	SI	+	+	+	-	+	+
V.escape 2	SI	+	+	+	-	+	+
V.escape 3	SI	+	+	-	-	+	+
V.escape 4	SI	+	+	-	-	+	+
E4	NSI	-	+	-	+	+	-
LN15	NSI	-	+	-	+	+	-
S8	NSI	-	+	-	+	+	-
B8	NSI	-	+	-	+	+	-
L5	NSI	-	+	-	+	+	-

*T-cell lines (C8166, Molt4 and SupT1), primary blood mononuclear T-cells (PBMC) and macrophages were infected (multiplicity of infection = 1) and the virus production (p24-ELISA) assessed.

†Transiently transfected CCC cat cells stably expressing human CD4 were treated with high-titre virus stocks (at least 1,000 times half-maximal tissue-culture infective dose for C8166 T-cells) and three days later fixed and immunostained for p24 expression and scored for foci of infection¹². + indicates that at least 100 infectious units of HIV-1 were recorded on transfected cells. -/+ indicates that use of CCR5 was one-tenth that of Gun-1 wild type. E4, LN15, S8, B8, L5: single-round infectivity assays with the envelope genes derived from different tissues were done using 10 ng p24 containing env-pseudotyped virus stock carrying a chloramphenicol-acetyl-transferase (CAT) gene in place of nef (kindly provided by J. Sodroski). Three days after infection the cells were lysed and the amount of CAT determined by ELISA (Boehringer). + indicates positive results in three independent experiments. SI, syncytium-inducing; NSI, non-syncytium-inducing.

- Weiss, R. A. & Clapham, P. R. *Nature* **381**, 647-648 (1996).
- Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. *Science* **272**, 872-877 (1996).
- Dragic, T. *et al.* *Nature* **381**, 667-673 (1996).
- Deng, H. *et al.* *Nature* **381**, 661-666 (1996).
- Zhang, L., Huang, Y., He, T., Cao, Y. & Ho, D. D. *Nature* **383**, 768 (1996).
- Doranz, B. J. *et al.* *Cell* **85**, 1149-1158 (1996).
- Simmons, G. *et al.* *J. Virol.* **70**, 8355-8360 (1996).
- Choe, H. *et al.* *Cell* **85**, 1135-1148 (1996).
- Connor, R. I., Paxton, W. A., Sheridan, K. E. & Koup, R. J. *J. Virol.* **70**, 8758-8764 (1996).
- Wu, L. *et al.* *Nature* **384**, 179-183 (1996).
- Trkola, A. *et al.* *Nature* **384**, 184-187 (1996).
- McKnight, A. *et al.* *J. Virol.* **69**, 3167-3170 (1995).
- Takeuchi, Y., Akutsu, M., Murayama, K., Shimizu, N. & Hoshino, H. *J. Virol.* **65**, 1710-1718 (1991).
- Donaldson, Y. *et al.* *Lancet* **343**, 382-385 (1994).
- McKnight, A. *et al.* *J. Virol.* (in the press).

The ascent of ideas in evolution

Monad to Man: The Concept of Progress in Evolutionary Biology

by Michael Ruse

Harvard University Press: 1997. Publication date: 13 February. Pp. 628. \$49.95, £33.50

David L. Hull

For the past generation or so, highly vocal critics of science have claimed that the same sorts of forces that influence all other human institutions influence science as well. Such factors as reason, argument and evidence that are supposed to be internal to science get all the attention in the old positivist literature, but what really matters in science, according to these critics, are such things as economic status, social class and gender. Michael Ruse has been anything but a major contributor to this movement. His early writings make him sound like a minimally reconstructed logical empiricist — not a positivist to be sure, but close. In this book, however, Ruse argues at some length that at least one factor external to science — a belief in progress — has strongly influenced what we now term evolutionary biology from the late eighteenth century to the present.

In *Monad to Man*, Ruse provides a history of evolutionary biology from its inception to its belated maturity that is full of insight. For scientist after scientist, he shows that the factors usually considered internal to science were far from sufficient to support any conclusions about how 'progressive' evolution has actually been. Yet nearly all scientists who had an opinion on the subject concluded that evolution is progressive. In the absence of factors internal to science sufficient for arriving at this conclusion, Ruse argues that a belief in progress enters into science as a cause, not an effect, as a premise, not a conclusion. With respect to progress and evolutionary biology, externalists are right.

In the Western world, three patterns of change have gained periodic popularity. The Greeks were fond of cycles, while the Bible encouraged a belief in decline and fall. 'Progress' did not become really popular until the late eighteenth and early nineteenth centuries. Intellectuals and the general public viewed everything from biological evolution to the course of society as on the rise. Although Ruse is concerned to show how widespread a belief in progress became, he mentions isolated departures as well, for example Buffon's notion of 'evolution' as degeneration, and Lyell's repeated cycles. One view of change that was not popular at the time, however, was the absence of pattern. Hints of evolution as a haphazard random walk can be found in

Darwin's writings, but evolution without direction did not come into its own until the past few decades.

Although Ruse is interested primarily in the extent of the belief in progress, he is forced to concentrate on major figures because evidence about them is available. He is able to present a good cross-section of major figures before 1859, from Buffon and Lamarck to Richard Owen and Darwin. After 1859 he is forced to reduce his attention even more drastically, concentrating on Anglo-American evolutionists. Ruse has ample justification for ignoring the French, as they remained peculiarly immune to the attractions of Darwinism. Because so few French scientists became evolutionists, the role of progress in their acceptance of evolution is moot. However, Ruse admits that omission of German evolutionists is as serious as it is pragmatically necessary. He simply cannot do them justice. But he conjectures that any study of German evolutionists would confirm his hypothesis even more strongly than does his study of Anglo-American evolutionists.

For each of the figures he discusses, Ruse addresses three issues: their beliefs about the history of life; the relation they saw

between science and society; and the extent to which their general conclusions outstrip the empirical data available to them. As Ruse interprets the historical data, evolutionary biologists viewed evolution as progressive on the basis of the most rudimentary, equivocal evidence and reasoned quite facetiously between science and society. That earlier workers let all sorts of considerations that we now take to be extraneous to science intrude upon their work is not likely to raise many hackles. Those were the bad old days. But Ruse pursues his topic into the present. Most readers will find his discussions of more recent evolutionary biologists — from Fisher, Haldane and Wright to W. D. Hamilton, G. C. Williams and Stephen Jay Gould — the most interesting and the most provocative part of his book. I know that I did.

Another intriguing part of Ruse's book is his extensive discussion of the efforts made by evolutionary biologists to transmute evolutionary biology into a mature discipline — mathematics and all. But did not Darwin and his immediate successors do that a long time ago? One of Ruse's insights is that they did not, and the chief villain was, of all people, T. H. Huxley. Huxley played a

Guidebook for the underwater eco-tourist

The 90,000 miles of shoreline around the United States give access to a diverse range of marine life. The sea life surrounding the barrier islands of North Carolina differs greatly from the creatures that inhabit the coral reefs around the Florida Keys. Some of the animals found in the clear, warm waters around the Hawaiian Islands also swim through the waters of the Pacific Northwest, but that may be the only similarity between the two habitats. Charles Seaborn takes the reader on a tour of this ecological diversity in *Underwater Wilderness: Life in America's National Marine Sanctuaries and Reserves* (Roberts Rinehart in cooperation with the Monterey Bay Aquarium, \$29.95, £19.99). The book could appeal to the eco-tourist as much as to the student of marine biology. Right: purple hydrocoral photographed off the coast of California.



central role in establishing science as a profession in Victorian England, but he did so only by excluding evolutionary biology. It was too hypothetical. In his popular presentations, Huxley was very much an evolutionist, but when he put on his professional hat he explicitly and pointedly omitted any reference to evolutionary theory.

Ruse's goal of discerning the influence of progress on the development of evolutionary biology is made more difficult by a gentlemen's agreement forged in the 1940s that the only way for evolutionary biologists to turn their subject into a mature discipline was to exclude any reference to such things as progress from their professional writings. Beliefs in progress can be found in the more popular writings of present-day and recently dead evolutionary biologists, for example Theodosius Dobzhansky, but, for most present-day biologists, Ruse is forced to read between the lines.

I was on Ruse's side until this juncture. He has done a massive amount of research on a topic that warrants careful investigation. However, when he turns to more modern biologists, he weaves his conceptual net so finely that almost no one can avoid being classed as a progressivist. For example, he classifies anyone who recognizes the hierarchical organization of life as a progressivist. He discerns a progressivist bias in Sewall Wright's adaptive landscape. Hence anyone who uses this metaphor is automatically an advocate of progress. Ruse sees glimmers of progress in G. C. Williams's dour worldview. He even thinks that Gould is a closet progressivist. According to Ruse, Gould opposes the idea of progress in biological evolution because he thinks that it threatens sociocultural progress.

By the time we get to the present, Ruse has made a strong case for the role of beliefs about progress in evolutionary biology. He detracts from his very strong general thesis by pushing too hard to make too many present-day evolutionary biologists progressivists. Certainly some do believe in evolutionary progress, but many do not. As I read excerpts from his interviews with present-day biologists, my pity for his subjects grew apace. He kept after them relentlessly to confess: "Were you or have you ever been a progressivist?" Finally, Niles Eldredge exploded: "I don't care! I really don't care!" Ruse does not have to go to such extremes. His book presents one long argument in support of the view that ideas of progress have been central to evolutionary biology up to the present. He does not have to force such workers as Williams and Gould into his mould, especially since his general conclusion is that today "there truly is a functioning, professional, evolutionary biology not permeated by ideas of progress". For some reason, ideas of progress are losing the force they once had.

Testing is a scientific virtue. Ruse supports his general thesis about progress by piling up positive instance upon positive instance, but, as he well knows, more convincing ways exist for providing inductive support. One is to look for apparent counter-instances. During the period that he investigates, physics was undergoing significant change. According to thermodynamics, the Universe is running down, to end eventually in heat death. If a belief in progress was so powerful at the time, then it must have presented a formidable barrier to the acceptance of this new theory. If it did, then Ruse's thesis receives even more support. If it did not, then he has to explain this difference in the reception of evolution and entropy.

What Ruse needs to do now is to go back and do for thermodynamics what he has done for evolutionary biology. But it would be churlish of me to suggest that, just as he has finished what surely must be termed a *magnum opus*, he must begin all over again. That is more than a body can bear. □

David L. Hull is in the Department of Philosophy, Northwestern University, Evanston, Illinois 60208-1315, USA.

Selling science

After the Breakthrough: The Emergence of High-Temperature Superconductivity as a Research Field

by Helga Nowotny and Ulrike Felt
Cambridge University Press: 1996. Pp. 210.
£35, \$49.95

John Ziman

The discovery of high-temperature superconductivity in 1986 was a moment of high promise for the physical sciences. It unexpectedly opened the door to an exciting new field of research. It challenged the theoretical conventional wisdom. It aroused expectations of untold technological marvels. It could easily be described to the general public. It was a romantic vindication of traditional, individualistic, 'small' science. It had every desirable feature of a milestone in the history of science.

But the past decade has not seen these diverse promises fulfilled. Levitating trains, compact efficient electric motors, zero-wastage power lines and what have you, failed to materialize. After the initial flurry of excitement, research on high-temperature superconductivity has settled down into a steady slog. Much has been found out, but nothing as astonishing as the original discovery. The theorists have rubbed each other up very vigorously, but mostly to produce smoke rather than light. The putative technological applications have scarcely reached the development stage, let

alone the market-place. Indeed, as Helga Nowotny and Ulrike Felt point out, the study of high-temperature superconductivity is still regarded as basic science, undertaken more by academic than by industrial researchers.

The authors' thorough analysis of the public presentation of high-temperature superconductivity in the press charts this return to 'normal' science. The glamour has faded. The media sportingly reported the succession of record-breaking critical temperatures, but their interest waned as this curve levelled off. Without fraud, social risk or indecorous public controversy, what would there be to report? From interviews with researchers and research managers in various countries, it is clear that this is not, after all, a field where a heroic lash-up of sealing wax, string, inspiration and perspiration can win out over the big battalions of elaborate instrumentation, wide-ranging intelligence operations and batteries of sophisticated techniques.

Even the expected breakdown of disciplinary boundaries has not been fully realized. The established experts in *low-temperature* superconductivity — mainly physicists — have not been swept aside by interdisciplinary task-forces of 'outsiders' from materials science and beyond.

The authors fully understand the science of high-temperature superconductivity, but do not go into it in detail. Their real interest, and the uniquely valuable theme of the book, is how policy was made. The discovery of the phenomenon affected the research system like an experimental probe. The transient response revealed how that system worked. Did it perform according to specification? How effectively could it exploit a sudden opportunity? Which structures proved flexible and adaptable, which hidebound and brittle? A Machiavellian sociologist could scarcely have devised a more instructive intervention.

The authors' first point is that talk about globalized industry and transnational science is premature. Basic research is still largely supported and shaped by national governments. But every country does this in a different way. Science policy-making is as diverse as 'policy', not as universal as 'science'. This diversity is apparent even across such culturally similar countries as Austria, Germany, the Netherlands and Switzerland, which the authors have studied at first hand. Their schematic accounts of how national research programmes in high-temperature superconductivity evolved in the United States, the United Kingdom and Japan indicate still wider variations in structure and outcomes.

Most working scientists would not usually be tempted to delve into a comparative study of the tangled relationships between research funding agencies, govern-

ment departments, academic institutions and individual research leaders in various countries. But the sober sociological style of this book is misleading. It is much more than a high-quality source of valuable — often chastening — historical material for policy practitioners and analysts trying to get to grips with the specific characteristics of national research systems. Its well-informed, dispassionate analysis of the reaction of science to the bombshell of high-temperature superconductivity arrives at a radical conclusion, which all scientists need to take very seriously indeed.

The authors see science as a traditional social institution going through a major transformation as it adapts to rapidly changing circumstances. Basic research is now carried out in an 'extended laboratory', open on every side to its political, economic and social environment. Individual intellectual creativity has to give way to collective political solidarity. Even mainstream scientists are being forced to use extra-scientific channels of communication to win friends, and are spending a great deal of their time 'scientizing' — carrying out an astonishing range of activities to exploit or ward off the multitude of forces that now shape every research scene.

The main reason is that 'technological promise' is becoming the main criterion of 'good' science. But this is a purely notional quality. When, for example, scientists claimed through the media that strong national support for research into high-temperature superconductivity would be the key to economic competitiveness, they knew that this was wildly exaggerated. But they felt their rhetoric was justified by the funds it would bring in. Egged on by the media, politicians accepted these claims, and then needed reassurance that they were wise to do so. So now they are all caught up in a credibility game where the bench researchers know harsh realities that are seldom publicly aired and that are carefully filtered out as progress reports rise up through the system to the politicians at the top. "Who sold what to whom?" ask the authors, but do not stay for an answer.

The authors care deeply for science and admire its institutional vitality and strength. They do not follow the current sociological fad of belittling scientific knowledge for being 'socially constructed'. But they show just how the political reconstruction of the research system is affecting the knowledge it produces. We all agree, don't we, that science is too serious a matter now to be left to the scientists? But this thoughtful book reminds us how close it is to becoming 'politics by other means'. □

John Ziman, emeritus professor of physics at the University of Bristol, is at 27 Little London Green, Oakley, Aylesbury, Buckinghamshire HP18 9QL, UK.

Nature versus nurture at 60 MHz

Last November an experiment of global proportions was set in motion. It involves thousands of people, yet no public money has been contributed, and the only equipment participants need is a personal computer running Windows 95 and a CD-ROM loaded with a computer game. *Creatures*, as the game is known, is touted as the "world's largest experiment in artificial life" (GT Interactive, £34.99, \$39.95). The creatures in question are norns — cute and inquisitive artificial 'life-forms', each equipped with its own neural net for a brain and its own artificial biochemistry, both genetically specified in its digital DNA. Norns are born, mate and then die in a virtual world known as Albia. *Creatures* has been turning heads in the scientific community. For the first time, fundamental features that constitute intelligent life have been simulated in a single system that can run in real time on a commercially available computer. The game has been modelled as closely as possible on biological systems. Norns have compartmentalized 'brains' mimicking the structure of human brains. They are born with some instincts, but almost all their behaviour is emergent, and much has been quite unexpected. Their ability to generalize, and so respond to new situations, can make them appear genuinely intelligent. Each norn hatchling has its own unique genome, and their 'owners' can e-mail them to their friends or find them on the World Wide Web (see <http://www.cyberlife.co.uk>). It is impossible to predict the outcome of this breeding experiment, but already norns with interesting and useful mutations have 'evolved'. Despite being marketed as 'artificial life', *Creatures* is more an innovation in



artificial intelligence. Although the game uses 'artificial life technology', such as artificial DNA and computational chemistry, this is simply a means to an end — to produce artificial intelligence. Ultimately for entertainment, the game makes inevitable compromises in its representation of life. Natural selection is abandoned in favour of direct interaction: owners must teach their norns to communicate, and guide them towards mates. But norns encounter few dangers in Albia. The underlying technology, *CyberLife*, is likely to have a greater impact than a computer game. Computer experts predict that it will change the way we interact with computers, as they respond more 'intelligently' to our needs. Artificial brains might allow neuroscientists access to a richer source of data than is available with techniques using real brains. But, in recognizing that we do not need to understand intelligence to create it, do we also concede that it is irreducible and that we may never understand in detail how the mind works? Should *CyberLife*'s developers succeed in creating "fully self-aware artificial life-forms", we will be forced to evaluate the status of such creatures.

Ayala Ochert is a freelance science journalist.

Play-safe politics

Scientific Uncertainty and Environmental Problem Solving

Edited by John Lemons

Blackwell Science: 1996. Pp. 433. £59.50, \$80

Timothy O'Riordan

The broadsheet press, always looking for new angles, is asking two questions of science in the late twentieth century. First, what is left to be discovered? The pace and scale of scientific effort are filling in the niches, but the true pioneers are few and far between. And, second, how far can science truly help society when evidence is genuinely ambiguous, traditional disciplinary models unsatisfactory, and policy-makers want justification for expensive courses of action that may disturb accepted social practices?

This excellent book addresses the second question. Lemons has brought together a fascinating, coherent set of essays on the strug-

gles within science when faced with indeterminacy in providing advice for environmental problem-solvers. With one exception (an Egyptian) the authors are all American, so the case studies tend to emphasize US traditions of problem-framing, mediation, litigation and regulation. These are all contentious arenas, where data are incomplete and outcomes unforecastable. Long-established procedures based on burden of proof, balance of the evidence, and credibility of witnesses no longer hold. What is at stake is the extension and possible reformulation of science itself in the face of the complex and interdisciplinary issues that environmentalism has shaped for politicians the world over.

Indeed, Lemons might have made more of this theme. Science is incomplete as an approach to handling coupled relationships between chaotic natural systems and equally (but differently) chaotic human interpretations and responses. Maybe there is a 'new' environmental science in the making — that of an interactive, adaptive, heuristic relation-

ship between the natural and social worlds that can be understood only as elements of a whole.

The opening essay by Kristin Shrader-Frechette admirably sets the scene for the book. Shrader-Frechette is a distinguished philosopher of science who here characterizes four types of scientific uncertainty: framing problems, modelling uncertainty, statistical uncertainty and decision-theoretic uncertainty. The first of these, framing problems according to predetermined values or institutional mandates, tends in practice to depend on only two parameters — falsification or provisional acceptance. So, where a 'no effect' hypothesis is accepted as the frame of reference, scientists will expect those wanting to argue that there will be an identifiable effect to demonstrate that they are more credible than those who argue for a lack of effect.

Shrader-Frechette argues that complex environmental issues generate a need for a third parameter: where there is insufficient evidence, or where the decision stakes are so high that a 'play safe' strategy is required, a decision, in some circumstances, should be weighted more towards ethics or presumed social consensus than is justified by the data.

In the chapters that follow, this third parameter is applied to biodiversity (Carpenter, Lemons), to water resources (Canter), to environmental impact assessment (Canter, El Sayed), to risk management (Perry and Vanderklein), and to the institutions of law (Brown) and policy (Caldwell). In all these cases, science engages with social values and weights; resonates with different stances on moral choice and equity considerations; and, above all, is adaptable to varying democratic structures. One way forward is through mediation-based techniques involving a more decentralized approach to this interactive science than the more traditional, top-down, consultative styles.

Turning to the second of the four types of uncertainty, modelling, Shrader-Frechette points out that all models are strong or weak in relation to their dependence on a series of nested assumptions (hypotheses) that must be internally consistent. The key is to use models that are robustly and openly tested against the evidence, or against various value positions. In practice this may not be possible, because the data are not reliable, or may not be done, because consultative processes are flawed or biased.

These points are given a good airing. Lyn-ton Caldwell, the creator of the US National Environmental Policy Act, points out that "conflict between traditional politics, conventional economics and scientific knowledge have increased with enactment of legislation to protect health and environment from hazards inherent in 'unecologic' aspects of human behavior". He blames the "pseudo science" of disguised uncertainty associated with statistical manipulation, and the over-expo-

Searching out Scotland's floral rarities

Scottish Wild Plants: Their History, Ecology and Conservation (The Stationery Office, £12.95)

focuses on more than 40 of the rarest and most interesting species in Scotland's flora. The text gives detailed factual material about the plants, including special features of interest, their ecological needs and status in the wild, and the history of their discovery and some of the colourful personalities involved. The authors are Philip Lusby, who runs the Scottish Rare Plant Project at the Royal Botanic Garden in Edinburgh, and ecologist Jenny Wright. The book is illustrated with fine photographs by Sidney J. Clarke, including this picture of purple saxifrage.



sure of more and more different possible outcomes from a given policy action because of the "niche-filling" science mentioned at the outset. The result, according to Caldwell, is a tendency in science for "prescriptive advocacy research" that defends a preset conclusion, often invoking models and predictions that cannot be fully justified.

The third main type of uncertainty, statistical uncertainty, comes in two forms, false positives and false negatives. Here the scientist tends to be more anxious to avoid suggesting a course of inaction that turns out to be wrong than a course of action that turns out to be unnecessary. Hence the bias is towards precaution and a dislike of suggesting possible effects when in reality there may be none.

Some of this type of thinking lay behind the early period of research into bovine spongiform encephalopathy, where there was insufficient evidence to prove no effect, but a fear among scientists of starting a public-health scare of potentially (and actually, as it turned out) huge economic proportions before the evidence was sufficiently conclusive. This is an example of why Shrader-Frechette's 'play-safe' parameter approach of negotiated science remains contested.

Decision-theoretic uncertainty tends to be used in two ways by statisticians and policy analysts. First is the maximin rule, based on maximizing the 'least-worst' outcome and deliberately avoiding the worst-case likelihood. This is the approach followed in many low-level toxicological studies (discussed here by Cairns and Smith) and in the dispersal of high-level nuclear waste.

The second approach is to maximize expected pay-off (or utility) based on bayesian principles of statistical analysis. The issue here lies in the handling of likely losers in

the pay-off valuation. The utilitarians assume that losers can be bought off while the gainers still win. The maximin is that players accept that the losers need special treatment so that the best solution is to give them that special treatment. This is why the approach is sometimes called the egalitarian strategy.

The book as a whole contains a balance of argument favouring egalitarianism over utilitarianism in the handling of decision rules for complex environmental issues. Indeed, this is probably the single most important credo it contains. This is a noble cause, and one that is politically pragmatic as well as more distributionally fair. In fact, there are plenty of cases where the utilitarian approach still holds sway, notably in the siting of unwanted facilities such as incinerators and toxic-waste clean-up centres.

This text, delightfully comprehensive as it is, nevertheless fails to look at the implications both for science and for politics of its findings. For science, there may well be the need for a more ethically and value-orientated perspective that is achievable only through openness to those who have a stake in the possible outcomes. Once different players see more clearly how their demands impose a host of 'costs' on others, there is a chance, in a political framework of negotiated adjustment, that the maximin rule can be put into effect. But it is not automatic. Uncertainty poses as great a challenge for decision structures and the decentralization of power as it does for science. The second part of this fascinating odyssey remains to be explored. □

Timothy O'Riordan is in the School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

Measurement of tectonic surface uplift rate in a young collisional mountain belt

Lon D. Abbott^{*†}, Eli A. Silver^{*}, Robert S. Anderson^{*}, Randall Smith[‡], James C. Ingle[§], Stanley A. Kling^{||}, David Haig[¶], Eric Small^{*}, Joseph Galewsky^{*} & William Sliter[#]

^{*} Department of Earth Sciences, University of California, Santa Cruz, California 95064, USA

[‡] GeoInformation Technology, 2771 E. Shaw Avenue, Fresno, California 93710, USA

[§] Department of Geological and Environmental Sciences, Stanford University, Stanford, California 94305, USA

^{||} 416 Shore View Lane, Leucadia, California 92024, USA

[¶] Department of Geology and Geophysics, University of Western Australia, Nedlands, Perth, Western Australia 6009, Australia

[#] United States Geological Survey, Western Geologic Mapping Team—Mail Stop 975, 345 Middlefield Road, Menlo Park, California 94025, USA

Measurement of the rate of tectonically driven surface uplift is crucial to a complete understanding of mountain building dynamics. The lack of a suitable rock record typically prevents determination of this quantity, but the unusual geology of Papua New Guinea's Finisterre mountains makes measurement of this rate possible. The tectonic surface uplift rate at the Finisterre range is 0.8–2.1 mm yr⁻¹, approximately that expected to arise from crustal thickening.

In the study of mountain belts, one of the most fundamental quantities that geoscientists seek to measure is the uplift rate. The term 'uplift' is often loosely applied and may refer to one of several distinct quantities¹. Among the most important types of uplift are: rock uplift, tectonic uplift at a point, exhumation, and tectonic surface uplift. Rock uplift is the total displacement of a point of rock with respect to sea level (or the geoid). Rock uplift can be driven by tectonic forces or by the isostatic response to erosion. Tectonic uplift at a point is that portion of rock uplift that is driven exclusively by tectonics. Exhumation is the sum of erosion and tectonic denudation. It tracks the displacement of rocks towards the Earth's surface, so, unlike rock uplift and tectonic uplift, which are referenced to sea level, exhumation measurements do not record elevation changes. Tectonic surface uplift is the mean displacement (relative to sea level or the geoid) of a broad region (over 1,000 km²) of the Earth's surface resulting solely from tectonic activity¹. Stated another way, tectonic surface uplift is the mean tectonic uplift for all points within an area larger than 1,000 km².

The importance of tectonic surface uplift lies in its utility for quantifying large-scale tectonic forces. Measurement of rock uplift or tectonic uplift at individual points yields information about local structures. However, because tectonic processes operate at crustal scales, it is average tectonic uplift over a broad area (tectonic surface uplift) that is needed to quantify the tectonic forces responsible for mountain building¹. Measurement of the tectonic surface uplift rate (TSUR) can help illuminate the mechanics of mountain building and provide important constraints on the behaviour of the continental lithosphere^{1,2}. For instance, although the main tectonic mechanism driving the growth of mountain ranges is thickening of the continental crust during plate collision³, several additional tectonic uplift mechanisms have been proposed^{2,4,5}. If supplementary mechanisms have contributed to unusually rapid uplift, the measured TSUR should exceed the rate expected from crustal thickening. Whether mountains can rise faster than is possible via crustal thickening is central to the recent debate over whether rapid uplift of Tibet caused late Cenozoic climate change^{6–8}. The TSUR also provides an important constraint for numerical models that explore dynamic feedbacks between lithospheric and surface processes during mountain building^{9,10}.

Despite its importance to modern quantitative tectonic studies, the TSUR is so difficult to quantify that it has never been successfully measured¹. Most uplift studies measure rock uplift (for example, geodetic surveys) or exhumation (for example, thermochronological studies). Valuable information is gleaned from these techniques, but because they do not measure TSUR, they cannot be used to quantify large-scale tectonic forces¹.

Here we show that the unusual geology of Papua New Guinea's Finisterre mountains (Fig. 1) allows quantification of the TSUR. Limestone plateaux, whose surfaces have suffered little or no erosion, provide uplift markers. Using biostratigraphic methods, we dated the surface rocks and determined their depositional water depth at 15 plateau-top control points. We then fitted a continuous surface to the plateaux and assigned each surface point an age and water depth by interpolating from the control points. Subtracting the present topography from the reconstructed surface yielded an erosion map, from which we calculated the isostatically driven rock uplift at each point that can be attributed to erosional unloading. The tectonic component of uplift for each point was derived by subtracting this isostatic uplift component from the total rock uplift. Finally, we calculated the TSUR by dividing the magnitude of tectonic uplift by the uplift duration at each point and averaging these values over all points on the surface.

The TSUR for the central Finisterre range is between 0.8 and 2.1 mm yr⁻¹. The uplift rate expected to be caused by crustal thickening is 0.6–3.0 mm yr⁻¹. The TSUR is therefore consistent with crustal thickening being the only cause of uplift, but the existence of additional mechanisms that either increase or retard uplift is not precluded. The Finisterre TSUR is also similar to the average rock uplift rate for Quaternary terraces along the range's northeastern coast^{11,12}. This similarity offers hope that data on terrace rock uplift may provide a proxy for the TSUR in some areas lacking the conditions necessary to measure the latter.

Measurement of the TSUR

The total uplift at a point, i , can be expressed as follows^{13,14} (Fig. 2):

$$U_i = (Z_i - Z_0) + E_i + \Delta H_{sl,i} \quad (1)$$

where, at any point i , U_i is the total uplift, Z_i is the present topographic elevation, Z_0 is the original topographic elevation

[†] Present address: Down To Earth Technology, 2964 Palo Parkway, Boulder, Colorado 80301, USA.

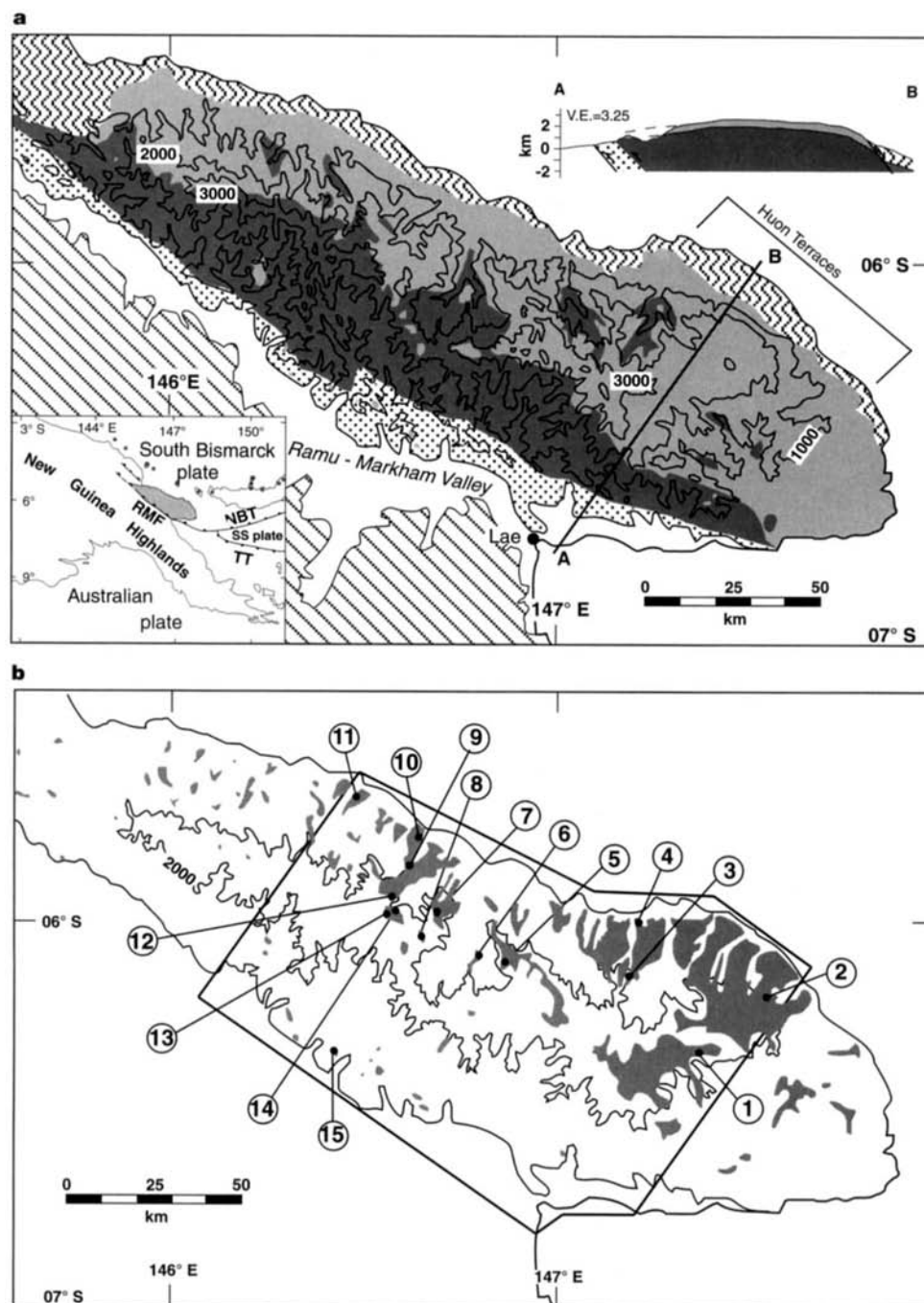


Figure 1 **a**, Simplified geological map of the Finisterre range with 1,000-m elevation contours. The Ramu-Markham Valley marks the collision suture. Patterns are as follows. Dark shading, Oligocene to Early Miocene volcanic and volcanoclastic rocks; light shading, Gowop Group marine carbonate rocks; stippling, Neogene and Quaternary clastic accretionary wedge; wave pattern, other Finisterre terrane rocks; hatching, New Guinea Highlands. Cross-section A-B illustrates the anticlinal nature of the range. The Finisterre range is shaded on the inset map showing the tectonic setting of the Finisterre arc-continent collision. RMF, Ramu-Markham fault (the collision suture); NBT, New Britain trench; SS plate, Solomon Sea plate; TT, Trobriand trough. **b**, Outcrops of Gowop Group rocks used to reconstruct the zero-erosion surface are shaded on this map of the Finisterre range. In places, the Gowop Group mapped in **a** is heavily

dissected. Only outcrops displaying little or no erosion, as evidenced by their plateau morphology and Nokopo Formation composition (the youngest unit of the Gowop Group), are used in the reconstruction and hence shaded here. A few of the smallest shaded areas are ridgetop outcrops of Gowop Group too small to show plateau morphology but they mark sites of modest or no erosion in otherwise heavily dissected areas. Plateau outlines were digitized from a digital elevation model that was constructed on a 55-m grid from 1:100,000 scale topographic maps using the TNTMips software. Numbers 1-15 mark the control points used in the TSUR calculation. Age and elevation data for these points are in Table 1. The 2,000-m elevation contour is shown. The bold outline surrounds the 12,000 km² region for which the TSUR calculation was done.

with respect to sea level at the time of formation, E_i is the thickness of eroded material, and ΔH_{sl_i} is the change in sea level (with a sea-level rise from the time of creation of the marker to the present assumed to be positive).

Uplift at a point can be driven by tectonics or by the isostatic response to exhumation. Some portions of any landscape are being eroded more rapidly than others, with erosion rate variations occurring over distances of metres to kilometres. In contrast, the isostatic response to erosion is regional; the lithosphere responds to the average amount of erosion over an area 10^4 km^2 or more¹⁵. Hence, areas that are eroding more slowly than the regional average will be displaced upwards (relative to sea level), even in the absence of tectonic forcing^{1,16}. The total uplift at a point can be expressed as the sum of its tectonic and isostatic components:

$$U_i = U_{ti} + U_{ei} \quad (2)$$

where, at the point i , U_{ti} is the tectonic component of uplift and U_{ei} is the erosionally induced isostatic component of uplift.

An expression for the tectonic component of uplift at point i can be derived^{13,14} by combining equations (1) and (2) and solving for U_{ti} :

$$U_{ti} = (Z_i - Z_{0i}) + E_i + \Delta H_{sl_i} - U_{ei} \quad (3)$$

The rate of tectonic uplift at point i is calculated by dividing this tectonic uplift expression by the duration of uplift, t_i :

$$\text{rate of tectonic uplift of point } i = \frac{U_{ti}}{t_i} \quad (4)$$

To be meaningfully applied to large-scale tectonic problems, the tectonic uplift rates calculated (from equation (4)) at all points on a surface greater than $1,000\text{--}10,000 \text{ km}^2$ must be averaged to obtain the TSUR¹:

$$\text{TSUR} = \frac{1}{N} \sum_{i=1}^N \left[\frac{(Z_i - Z_{0i}) + E_i + \Delta H_{sl_i} - U_{ei}}{t_i} \right] \quad (5)$$

where N is the number of sample points.

Examination of equation (5) makes it clear that measurement of the TSUR requires the following: (1) Preservation of a datable rock sequence that was deposited during or immediately before uplift (to constrain t_i); (2) a record of depositional elevation for that rock sequence (to constrain Z_{0i}); (3) measurement of the erosional thickness (E) at every point and knowledge of the original density of the missing material (to calculate U_{ei}); and (4) knowledge of the change in sea level between the time of deposition and the present.

The geological record in most areas is too incomplete to satisfy all the above constraints. The syn-orogenic sediments that meet requirement (1) are usually quickly eroded. On high plateaux these sediments are often preserved, but most are deposited above sea level, making it difficult to determine their depositional elevation (requirement (2)). Finally, it is rare to find surfaces experiencing no erosion within mountain ranges. Without such reference points, the total amount of eroded material (requirement (3)) is difficult to quantify.

Application to the Finisterre range

The Finisterre mountains formed 3.0–3.7 Myr ago in response to collision between the Finisterre volcanic arc terrane and the Australian continental margin¹⁷. Oligocene and Early Miocene volcanic and volcanoclastic rocks of the Finisterre arc are overlain unconformably by a sequence of Neogene and Quaternary marine carbonate rocks (the Gowop Group)¹⁸. This sequence is broadly folded and thrust over clastic rocks of a Neogene accretionary wedge (Fig. 1)^{19,20}. The broad anticline forming the range has been eroded into a series of high plateaux separated by deep gorges.

The Finisterre range possesses all of the attributes necessary for

Table 1 Data for each control point on the plateau surface

Control point	Age (Myr)	Original water depth (m)	Modern elevation (m)	ΔH_{sl} (m)	U_i (mm yr ⁻¹)
1	3.0–3.1	500–2,000	2,460	–60 to +70	0.9–1.5
2	0.95–1.7	1,600–2,000	1,870	–20 to +90	2.0–4.2
3	2.55–2.7	2,000–3,000	2,040	–20 to +90	1.5–2.0
4	1.0	2,000–3,000	210	–20 to +85	2.2–3.3
5	3.0–5.2	150–3,000	2,400	–90 to +95	0.4–1.8
6	4.0	1,200–2,000	1,840	–70 to +55	0.7–1.0
7	1.7–1.9	150–500	2,530	–20 to +90	1.4–1.8
8	3.0–3.8	500–2,500	2,470	–70 to +80	0.8–1.7
9	3.6–4.2	150–3,000	1,500	–90 to +80	0.4–1.3
10	3.0–3.1	150–3,000	410	–60 to +70	0.2–1.2
11	0.95–1.7	50–100	140	–20 to +90	0.1–0.4
12	2.7	1,500–3,000	2,500	+20 to +90	1.5–2.1
13	1.1–1.8	0–150	2,620	–20 to +90	1.4–2.6
14	2.4–2.6	1,200–2,500	2,450	–20 to +100	1.4–2.1
15	1.0–2.0	500–2,000	680	–20 to +90	0.6–2.8

Control point locations are shown in Fig. 1b. Age ranges are based on palaeontological zonations calibrated to the most recent timescale of Berggren *et al.*^{36,37}. The uncertainty in the age determined for a given sample is a function of the preservation and diversity of its microfossil biota and the stratigraphic range of that biota. Samples listing a single age either contain a biota diagnostic of a well-dated biostratigraphic event or are assigned an age by extrapolation from well-dated samples lower in the section by assuming a constant rate of sediment accumulation. Original water depths are calibrated using the classification of Ingle³⁸. Modern elevation is the present sample elevation, ΔH_{sl} is the sea level change from the time of deposition to the present³⁹, and U_i is the rock uplift rate at each control point.

determination of the TSUR. The plateaux are capped by marls deposited during collision whose age and original depth of deposition can be determined (requirements (1) and (2)), and the plateau tops are zero-erosion surfaces that provide reference points for the thickness of eroded material in the gorges (requirement (3)). Finally, ΔH_{sl_i} can be obtained from eustatic sea-level curves (requirement (4)).

The age of the plateaux. Marls of the Nokopo Formation, the youngest unit of the Gowop Group¹⁸, cap most of the plateaux. The marls contain benthic and planktonic foraminifera and calcareous nannoplankton that we dated using biostratigraphic methods. We examined 207 Nokopo Formation samples from 20 sites throughout the central Finisterre range. Only samples from the formation's stratigraphic top were used to constrain plateau ages. Samples from 15 sites yielded useful age information (Fig. 1b, Tables 1 and 2).

Determination of any post-depositional erosion of the plateau surfaces is crucial. If erosion has occurred, the values used for E_i and U_{ei} in equation (5) will be too low and those used for t_i will be too high, making the calculated TSUR a minimum. Several lines of evidence suggest that erosion has been minimal. (1) Porosities in plateau-top Nokopo Formation outcrops are commonly $\geq 50\%$ (L.D.A., unpublished data); such high porosities would not be maintained if the samples have been buried more than a few tens of metres (ref. 21). (2) The soft, easily erodable Nokopo Formation caps most plateaux that we visited. As in most areas the formation is 50–100 m thick¹⁸, only a small amount of erosion would strip the unit from the plateaux. (3) Modern plateau erosion is dominated by the formation of karst dolines, leaving much of the surface intact. (4) The gentle dips of most Nokopo Formation outcrops are concordant with the dip of the plateau surface; erosional plateaux often display discordance. (5) A low vitrinite reflectance (R_0) value of 0.15% was obtained near the base of an anomalously thick (623 m) Nokopo Formation section, indicating that the sample has not been exposed to temperatures $>65^\circ\text{C}$ (M. Bustin, personal communication). Such extreme thermal immaturity precludes substantial erosion²².

Examination of Table 1 shows that the plateau surface ages differ

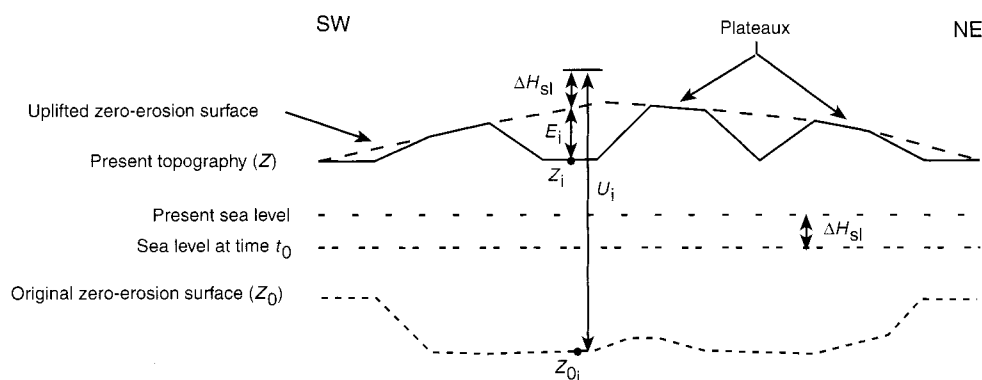


Figure 2 Schematic cross-section of Finisterre range uplift illustrating the quantities used in the TSUR calculation. A depositional surface originally below sea level (Z_0 , short dashed line) was uplifted and eroded to form the present topography (Z , solid line). U_i is the total uplift at point i . Plateaux preserve part of

the original surface, but erosion (E_i) has formed canyons elsewhere. The long dashed line connecting the plateaux reconstructs a surface of zero erosion. ΔH_{sl} is the change in sea level between time t_0 , when surface Z_0 was formed, and today.

by as much as three to four million years. It is possible that this diachrony indicates erosion at some control points; however, diachrony of the zero-erosion surface is expected because uplift of the range has progressed from northwest to southeast^{17,23} and because the range has been uplifted as a broad dome rather than uniformly. Turbidite sediments would bypass the dome's rapidly rising uplift axis sooner than they would the more gently rising dome flanks, causing diachrony of the zero-erosion surface and helping to explain both the variable thickness of the Nokopo Formation and the lack of accumulation of younger, shallower-water sediments later in the submarine uplift phase. Bottom currents may also have prevented accumulation of younger hemipelagic sediments. The ages of plateau-top samples are consistent with the direction and timing of collision propagation^{17,23}, and with the expectation of younger ages on the dome's flanks.

The original elevation of the plateau-top marls. The original depositional water depth for each Nokopo Formation sample was deduced from benthic foraminifera. Most of the benthic foraminifera identified in our samples are still living in the southwestern Pacific Ocean and their established modern depth ranges in this region were used to interpret original depth of deposition. The depth ranges shown in Table 1 reflect the maximum and minimum possible depths of deposition. The species used in the depth determination for a representative sample are shown in Table 2.

Determination of the depositional water depth produces the single largest source of uncertainty in the TSUR calculation because most of the Nokopo Formation was deposited at lower bathyal to lower middle bathyal oceanic depths, the region including the broadest depth ranges exhibited by modern species (1,500–4,000 m depth). However, specimens of deeper dwelling species in

Table 2 Microfossils identified at control point 2

Calcareous nannoplankton	Planktonic foraminifera	Benthic foraminifera
<i>Calcidiscus macintyreii</i> (F)	<i>Candeina nitida</i>	<i>Cassidulina</i> sp.
<i>Cyclargolithus floridanus</i> (R)	<i>Globigerinella aequilateralis</i>	<i>Cassidulina crassa</i>
<i>Dictyococcites antarcticus</i> (V)	<i>Globigerinoides ruber</i>	<i>Ceratobulimina pacifica</i>
<i>Dictyococcites minutus</i> (A)	<i>Globigerinoides conglobatus</i>	<i>Cibicidoides bradyi</i>
<i>Dictyococcites productus</i> (F)	<i>Globigerinoides sacculifer</i>	<i>Fissurina</i> sp.
<i>Discoaster asymmetricus</i> (R)	<i>Globorotalia hirsuta</i>	<i>Globocassidulina subglobosa</i> †
<i>Discoaster brouweri</i> (R)	<i>Globorotalia menardii</i>	<i>Hoeglundina elegans</i>
<i>Discoaster deflandrei</i> (V)	<i>Globorotalia scitula</i>	<i>Laticarinina pauperata</i>
<i>Discoaster surculus</i> (R)	<i>Globorotalia tumida</i>	<i>Melonis pompilioides</i> †
<i>Emiliania annula</i> (F)*	<i>Globorotalia tosaensis tenuithec</i> *	<i>Nodosaria</i> sp.
<i>Gephyrocapsa (small)</i> (C)*	<i>Globorotalia truncatulinoides</i> *	<i>Oridorsalis tener</i>
<i>Helicosphaera carteri</i> (C)	<i>Neogloboquadrina dutertrei</i>	<i>Oridorsalis umbonatus</i> †
<i>Helicosphaera sellii</i> (F)*	<i>Orbulina universa</i>	<i>Planulina wuellerstorfi</i> †
<i>Pontosphaera</i> spp. (V)	<i>Pulleniatina obliquiloculata</i> *	<i>Pleurostomella</i> sp.
<i>Reticulofenestra minuta</i> (A)		<i>Pyrgo murrhina</i>
<i>Reticulofenestra minutula</i> (C)		<i>Quinqueloculina</i> sp.
<i>Reticulofenestra pseudoumbilica</i> (R)		<i>Sigmilopsis schlumbergi</i>
<i>Sphenolithus moriformis</i> (F)		<i>Stilostomella</i> sp.
<i>Thoracosphaera</i> sp. (R)		<i>Uvigerina hispida</i> †
		<i>Valvulinera</i> sp.

* Key planktonic foraminifera and nannoplankton species for age determination.

† Key benthic species for estimating depth of deposition.

V, very rare (1 specimen); R, rare (2–10 specimens); F, few (11–100); C, common (101–1,000); A, abundant (>1,000).

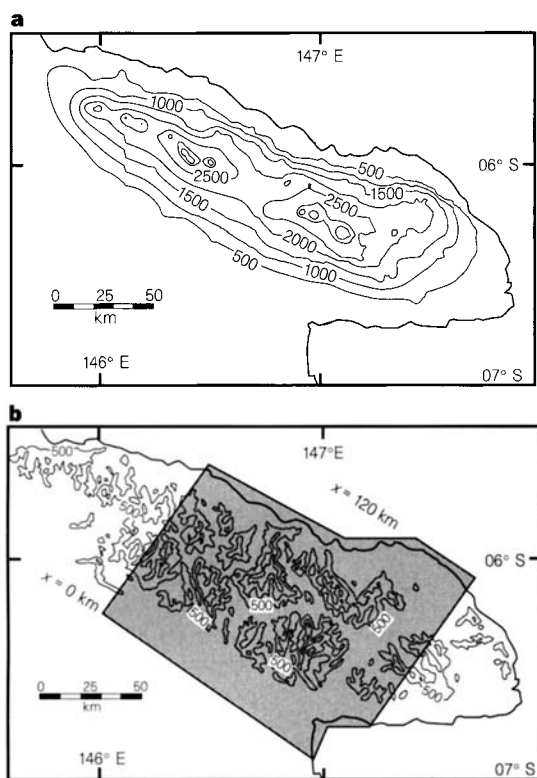


Figure 3 **a**, Contour map of the zero-erosion surface. The plateaux shown in Fig. 1b were imported to the Geographical Resource Analysis Support System (GRASS) software, where a regularized spline was fitted to them. A high tension of 300 was used to produce a surface that stretches tautly between extant plateaux^{24,25}. Contour interval is 500-m. **b**, Map of erosion height with 500-m contour interval. The map was generated by subtracting the existing topography (Fig. 1a) from the reconstructed depositional surface (**a**). The shaded area shows the TSUR calculation space, which is also the region used to produce the mean erosional profile of Fig. 4a. The $X = 0$ line marks the collision suture in the Ramu-Markham Valley. The Australian plate may be broken at the $X = 120$ km line²⁶.

control samples lack the corrosion that is common to foraminifera deposited at or approaching the carbonate compensation depth, restricting their maximum water depth to $<3,000$ m.

A much smaller uncertainty arises in determination of the present sample elevation. Elevation was determined using barometric altimetry and by locating samples on 1:100,000 scale topographic maps with 40-m contour intervals. The average elevation uncertainty is ± 35 m.

Reconstruction of the zero-erosion surface. Finisterre plateaux on opposite sides of deep river gorges lie at nearly identical elevations, indicating that they are remnants of an originally more extensive surface. We reconstructed this zero-erosion surface by generating a digital elevation model, on which we digitized the outlines of the existing plateaux (Fig. 1b). Fitting a continuous surface to the plateaux (Figs 2, 3a) was done using a regularized spline with adjustable tension that can handle irregularly spaced data^{24,25}. By subtracting the present topography from the reconstructed surface, we produced an erosional thickness (E_i) map (Fig. 3b), which is needed to calculate the magnitude of erosionally induced uplift (U_e).

Mistakes in reconstruction of the zero-erosion surface would

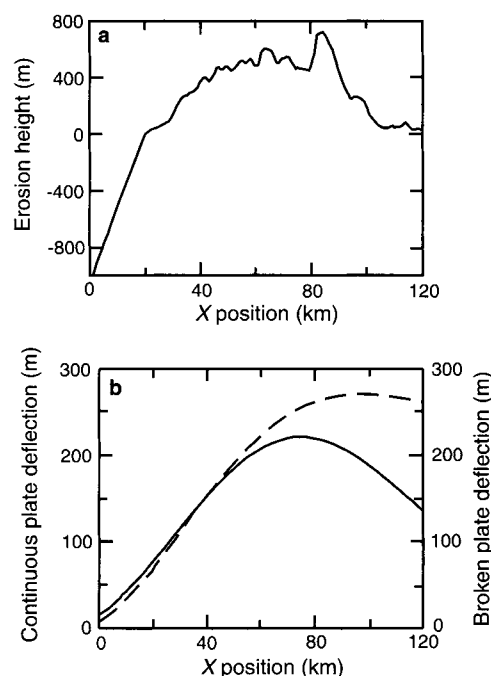


Figure 4 **a**, Profile of erosion thickness across the range (from $X = 0$ to $X = 120$ km, see Fig. 3b), constructed by averaging the erosion height along a swath 140 km wide (Fig. 3b) parallel to the strike of the range. Deposition in the Ramu-Markham Valley (Fig. 1a) also causes flexure, and is shown as a negative erosion height. The valley's sedimentary section thickens southwestwards from 300 m near the range front to about 1,000 m near the southern edge³⁹. **b**, Profiles of erosionally driven uplift (U_e) at each point due to the erosional load profile shown in **a**, for both a continuous plate (solid line) and a plate broken at $X = 120$ km (dashed line, see Fig. 3b). The calculations follow Small and Anderson¹⁶ and assume: effective elastic thickness = 20 km; Poisson's ratio = 0.25; Young's modulus = 7×10^{10} Pa; flexural rigidity = 5×10^{22} N m. The flexural model is a one-dimensional approximation for the response to a two-dimensional erosion pattern (Fig. 3b). A test of the calculation's sensitivity to this approximation is to vary the swath width along which erosion is averaged in **a**. If the one-dimensional approximation is insufficient, then the U_e values obtained using a narrow swath will differ significantly from those obtained using a wide swath. The maximum difference in U_e at any single point for swath widths between 10 km and 190 km is small (± 25 m for a continuous plate and ± 38 m for a broken plate), justifying the one-dimensional approximation.

cause errors in the erosion thickness map and therefore in the erosionally driven uplift component (U_e). Although we cannot assign quantitative uncertainties to the surface misfit, we believe it to be small. The reconstruction (Fig. 3a) provides a good fit to the existing plateau on the Huon peninsula, where it forms a nearly intact, broad arch (Fig. 1). Plateau surfaces are extensive and closely spaced on the northeast flank of the range (Fig. 1b), minimizing fitting errors here. Plateaux are rare along the southwest flank of the range, but a few small, widely distributed outcrops of the Gowop Group, which immediately underlies the plateaux elsewhere, strongly constrain acceptable surface fits. The broad anticline forming the Finisterre range is disrupted by smaller folds and faults²⁰. However, these are low-amplitude, second-order complications that do not cause significant errors in the surface reconstruction.

The magnitude of erosional uplift. Knowing the distribution of erosion (Fig. 3b), it is possible to calculate the magnitude of erosionally induced uplift (U_e) at each point. The flexural rigidity of the lithosphere in the Finisterre collision zone is between 1×10^{22} and 2×10^{23} N m (refs 26, 27). At such rigidities, isostatic compensation of the 80–120 km wide Finisterre range will be regional.

Using the method of Hetenyi²⁸, we constructed a one-dimensional elastic flexure model to calculate the deflection field (U_e) that would result from the Finisterre erosion pattern. The erosional load used in the model is a range-normal profile of average erosional thickness across a swath 140 km wide (Fig. 4a).

Figure 4b depicts the flexural response (U_e) to the erosional load profile of Fig. 4a. The Australian plate may be broken near the northern edge of the Finisterre range (Fig. 3b)²⁶, but an enigmatic plane of seismicity at about 100 km depth leaves the mechanical nature of the lithosphere under the Finisterre range uncertain²⁹. Given this uncertainty, we performed U_e calculations for both continuous and broken plate models, assuming an intermediate flexural rigidity of $D = 5 \times 10^{22}$ N m. Using the published end-member values for D produces a maximum uncertainty in U_e of ± 85 m for a continuous plate and ± 75 m for a broken plate. Our one-dimensional approximation results in only small uncertainties because the lithosphere responds to regional erosion, reducing the effect of local erosional thickness variations (Fig. 4b).

Calculation of palaeo sea level. Conflicting eustatic sea-level curves have been published for the time period (1–4 Myr ago) of Nokopo Formation deposition, and rapid (10^4 – 10^5 yr) oscillations in sea level that have occurred since the Pliocene are not accounted for in all curves³⁰. We used the endmember values of published sea-level curves^{30–32} for the ΔH_s range at each control point (Table 1).

Calculation of the tectonic surface uplift rate. We gridded the digital elevation model at 1-km intervals and assigned every grid point values for each parameter in equation (5). This involved interpolation of age and depositional depth data between control points. It is difficult to quantify the uncertainties associated with this interpolation, but abrupt age discrepancies (apart from the systematic diachrony discussed above) are unlikely, given that the tectonic and oceanographic controls on deposition would be similar throughout the area. It is also unlikely that major bathymetric ridges with enough relief to affect our calculations have gone undetected, given the control point spacing.

To calculate the TSUR, we applied equation (5) to all gridded points in a 12,000 km²-calculation space that includes both remnants of the zero-erosion surface and areas dissected by erosion (Figs 1b, 3b). The calculation space is large enough to provide a meaningful TSUR¹. The calculation was run for both continuous and broken plate models. The maximum possible TSUR was calculated using the youngest sample age, deepest original depth, lowest palaeo sea level, and smallest erosional uplift component. The minimum possible TSUR was calculated using the antithetic endmember values for each parameter. The TSUR is 0.86–2.06 mm yr⁻¹ for a continuous plate and 0.84–2.03 mm yr⁻¹ if the plate is broken.

All quantifiable uncertainties have been included in the calculations. The most significant of these are the depositional depth, which can be as much as 2,950 m for a given sample, and the sample age, which can be as great as 2.2 Myr. Provided that the plateau surfaces do represent surfaces of zero erosion and that our reconstruction of the original zero-erosion surface is accurate, uncertainties stemming from all other aspects of the calculation are relatively small (at most 432 m for a given point). The TSUR is clearly not sensitive to whether the Finisterre lithosphere behaves as a continuous or a broken plate.

Implications of the calculated TSUR

Because our study is the first to calculate the TSUR, it allows the first comparisons to be made between that rate and other measures of uplift. Rock uplift rates of 2.0–7.6 mm yr⁻¹ have been measured over the past 21 kyr at several locations in the Ramu–Markham Valley (Fig. 1a), the collision suture zone³³. A suite of coastal coral terraces on the Huon peninsula (Fig. 1a) display rock uplift rates over the past 340 kyr decreasing from 3.0 to 0.5 mm yr⁻¹ away from the suture zone^{11,12}. The average rock uplift rate for the terraces,

derived from a contour map of uplift rates¹¹, is approximately 1.5 mm yr⁻¹, similar to the 0.8–2.1 mm yr⁻¹ TSUR. These results suggest that, although individual rock uplift rates are not indicative of the TSUR, the average rock uplift rate for the terraces does offer a reasonable approximation of the TSUR. Quaternary coastal terraces are the only extensive, datable surfaces in many mountain ranges. Although our results pertain only to the Finisterre range, the similarity between the average terrace rock uplift rate and the TSUR suggests that where rock uplift rates can be averaged across extensive constructional coastal terrace suites, this rate may provide a reasonable proxy for the TSUR.

The method used here for measuring the TSUR illustrates the many issues that must be addressed if tectonically meaningful uplift rates are to be determined. These issues are particularly important for models that attempt to unravel dynamic feedbacks between lithospheric and surface processes during mountain building^{9,10}. These feedbacks can be misrepresented if measured rock uplift rates are used instead of the TSUR; the rock uplift rates are themselves manifestations of these feedbacks because they result from a combination of both tectonic and erosional processes.

Knowing the TSUR, we may determine whether crustal thickening explains the rate of rise. Uplift caused by crustal thickening can be estimated using either an isostatic balance between normal and crustally thickened lithosphere, or by modelling the range as a fault bend fold formed over a crustal-scale fault ramp. Both methods require some input parameters that are poorly constrained for the Finisterre collision.

The first method uses an equation for the surface elevation difference (Δe) between crustally thickened and normal columns of lithosphere²:

$$\Delta e = (S - S_0) \left(\frac{\rho_{m0} - \rho_{c0}}{\rho_a} \right) - \frac{\alpha T_1}{2L} \\ \times \left[\left(\frac{\rho_{m0} - \rho_{c0}}{\rho_a} \right) \left(\frac{S^2}{\gamma} - S_0^2 \right) + \frac{\rho_{m0}}{\rho_a} (\gamma - 1) L^2 \right]$$

where ρ_{m0} is the 0 °C density of the mantle (3,330 kg m⁻³), α is the volume coefficient of thermal expansion (3.4×10^{-5} K⁻¹), T_1 is the temperature at base of lithosphere (1,333 °C), S is the present crustal thickness, S_0 is the crustal thickness before collision, L is lithospheric thickness in an unstrained state, ρ_a is the asthenosphere density = $\rho_{m0}[1 - \alpha T_1] = 3,179$ kg m⁻³, ρ_{c0} is the crustal density, and $\gamma = S/S_0$, the crustal thickening factor. The values cited for ρ_{m0} , α and T_1 are taken from England and Houseman². The crustal thickness of the Finisterre collision zone is not well constrained, but L.D.A. *et al.*²⁰ estimated $S = 52$ km and $S_0 = 37$ km. As the lithospheric thickness has not been measured, we performed calculations for a span of reasonable values ($L = 80$ – 120 km). Following England and Houseman², the crustal density was chosen such that the initial continental column is in isostatic balance with a standard mid-ocean ridge column, but we modified the initial continental column used by those authors to consist of between 775 and 2,115 m of water overlying continental crust. This initial column reflects conditions in the Finisterre area before collision and results in $\rho_{c0} = 2,692$ – $2,855$ kg m⁻³. For lithospheric thicknesses between 80 and 120 km, Δe is in the range 1,450–1,840 m. Uplift began in the southwestern portion of the study area about 2.5–3.0 Myr ago and in the northeastern portion about 1.0–1.5 Myr ago^{17,23}. Dividing these elevation change values by the duration of uplift in each area yields average uplift rates due to crustal thickening of 0.6–1.3 mm yr⁻¹.

Alternatively, the range can be modelled as a fault bend fold³. The Finisterre range is an anticline formed above a terrane-bounding fault^{19,20} that appears from earthquake data to dip $\sim 15^\circ$ NE over most of its length and steepen near the suture zone^{26,34}. If the convergence rate and ramp angle are known, then trigonometry can

be used to calculate the crustal thickening rate over the ramp. The advantage of this approach is that crustal and lithospheric thicknesses do not have to be specified; a drawback is that, because uplift rates are low in areas not moving up the ramp³, the portion of the study area ascending the ramp at any given time must be estimated.

South Bismarck–Australian plate convergence is 82 mm yr⁻¹ in the vicinity of the study area³⁵. As much as 45 mm yr⁻¹ of this convergence may be accommodated by deformation within the Australian plate itself²⁶, making the convergence rate across the Finisterre mountains 37–82 mm yr⁻¹. The rate of crustal thickening over a 15° fault ramp expected for these convergence rates is 9.6–21.2 mm yr⁻¹. Accounting for isostatic compensation, this crustal thickening rate would yield an uplift rate of 1.7–3.8 mm yr⁻¹ over the fault ramp if the entire range were ascending the fault ramp throughout the collision. Given the convergence and collision propagation rates, it is more likely that the northeastern portion of the 80–120 km wide range reached the fault ramp 1–2 Myr after the leading (southwestern) edge of the range. Therefore, on average, each point in the study area was ascending the ramp for roughly 50–80% of the collision duration, resulting in an expected TSUR of 0.9–3.0 mm yr⁻¹.

The measured TSUR (0.8–2.1 mm yr⁻¹) and the expected uplift

rates using both methods described above (0.6–1.3 mm yr⁻¹ and 0.9–3.0 mm yr⁻¹) all agree over most of their respective ranges, supporting a conclusion that the Finisterre range is built solely by crustal thickening processes. However, the range in these data allows for the TSUR either to exceed the uplift rate expected due to crustal thickening, in which case additional uplift mechanisms must be sought^{2,4,5}, or to fall below the crustal thickening uplift rate. An unexpectedly slow TSUR could be caused by subsurface loads beneath the range³, the presence of which is suggested by gravity data²⁶.

The TSUR provides an important constraint on subsurface processes that is not available for any other mountain range. Unfortunately, other critical data for the Finisterre range are currently lacking, making it difficult to determine whether the observed TSUR is the exclusive product of crustal thickening or if other processes are contributing to growth of the range. New studies are needed to better constrain such parameters as the range's crustal and lithospheric thicknesses and the convergence rate across the range. Used in conjunction with such new data, the TSUR could quantify the rates and magnitudes of the lithospheric processes responsible for range formation more fully than has been possible elsewhere. □

Received 17 January; accepted 18 December 1996.

- England, P. & Molnar, P. *Geology* **18**, 1173–1177 (1990).
- England, P. C. & Houseman, G. A. *J. Geophys. Res.* **94**, 17561–17579 (1989).
- Molnar, P. & Lyon-Caen, H. in *Geol. Soc. Am. Spec. Publ.* **218**, 179–207 (1988).
- Bird, P. *J. Geophys. Res.* **84**, 7561–7571 (1979).
- Charlton, T. R. *Geology* **19**, 28–31 (1991).
- Raymo, M. E., Ruddiman, W. F. & Froelich, P. N. *Geology* **16**, 649–653 (1988).
- Molnar, P. & England, P. *Nature* **346**, 29–34 (1990).
- Molnar, P., England, P. & Martinod, J. *Rev. Geophys.* **31**, 357–396 (1993).
- Koons, P. O. *Am. J. Sci.* **289**, 1041–1069 (1989).
- Beaumont, C., Fullsack, P. & Hamilton, J. in *Thrust Tectonics* (ed. McClay, K. R.) 1–18 (Chapman & Hall, 1992).
- Chappell, J. *Geol. Soc. Am. Bull.* **85**, 553–570 (1974).
- Chappell, J. *Search* **14**, 99–101 (1983).
- Garfunkel, Z. *Tectonophysics* **150**, 33–49 (1988).
- Brown, R. W. *Geology* **19**, 74–77 (1991).
- Tsuboi, C. *Gravity* (George Allen & Unwin, 1983).
- Small, E. E. & Anderson, R. S. *Science* **270**, 277–280 (1995).
- Abbott, L. D. *et al.* *J. Sed. Res.* **B64**, 169–183 (1994).
- Abbott, L. D. in *MARKHAM 1:250,000 Geological Series (Revision 1): Explanatory Notes* (eds Findlay, R. H., Arumba, J., Kopy, G. & Nekitel, S.) (Geological Survey of Papua New Guinea, in the press).
- Davies, H. L. *et al.* *Geo-Marine Lett.* **7**, 143–152 (1987).
- Abbott, L. D., Silver, E. A. & Galewsky, J. *Tectonics* **13**, 1007–1034 (1994).
- Angevine, C. L., Heller, P. L. & Paola, C. *AAPG Short Course Note Series No. 32* (American Association Petroleum Geologists, 1990).

- Katz, B. J., Pfeifer, R. N. & Schunk, D. J. *Am. Assoc. Petrol. Geol. Bull.* **72**, 926–931 (1988).
- Gill, J. B., Morris, J. D. & Johnson, R. W. *Geochim. Cosmochim. Acta* **57**, 4269–4283 (1993).
- Mitasova, H. & Mitas, L. *Math. Geol.* **25**, 641–655 (1993).
- Mitasova, H. & Hofierka, J. *Math. Geol.* **25**, 657–669 (1993).
- Abers, G. A. & McCaffrey, R. *Tectonics* **13**, 227–245 (1994).
- Galewsky, J., Silver, E. A. & Abbott, L. D. *Eos* **73**, 535 (1992).
- Hetenyi, M. *Beams on Elastic Foundation* (University of Michigan Press, Michigan, 1946).
- Abers, G. A. & Roecker, S. W. *J. Geophys. Res.* **96**, 6379–6401 (1991).
- Quinn, T. M. & Matthews, R. K. *Geology* **18**, 942–945 (1990).
- Bag, H. U., Hardenbol, J. & Vail, P. R. in *SEPM Spec. Publ.* **42** (ed. Wilgus, C. K.) 71–108 (Soc. Economic Paleo. Mineral., 1988).
- Prentice, M. L. & Matthews, R. K. *Geology* **16**, 963–966 (1988).
- Crook, K. A. W. *Tectonophysics* **163**, 105–118 (1989).
- Kulig, C., McCaffrey, R., Abers, G. A. & Letz, H. *Tectonophysics* **227**, 81–93 (1993).
- Taylor, B. *Geology* **7**, 171–174 (1979).
- Berggren, W. A. *et al.* *SEPM Spec. Publ.* **54**, 129–212 (1995).
- Berggren, W. A. *et al.* *Geol. Soc. Am. Bull.* **107**, 1272–1287 (1995).
- Ingle, J. C. *Cushman Foundation Spec. Publ.* **19**, 163–195 (1980).
- Pettifer, G. R. *Austr. BMR Rec.* **174/65** (1974).

Acknowledgements. J. Werlin, S. Nekitel, Hon. G. G. Saonu (MP) and many residents of the Finisterre range assisted with sample collection. The manuscript was improved by reviews from D. Burbank and T. Cook. The project was supported by grants from the Earth Sciences Program of the National Science Foundation (E.A.S. and R.S.A.).

Correspondence should be addressed to L.D.A. (e-mail: lon@earthsci.ucsc.edu).

KNOW YOUR COPYRIGHT RIGHTS RESPECT OURS

The publication you are reading is protected by copyright law.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931 Fax: 0171 436 3986

Notes for contributors

Nature is a weekly international journal covering all the sciences. Its readership is interdisciplinary, so manuscripts should be written clearly and simply, authors remembering that English is not the first language of many readers. Competition for *Nature's* limited space is severe, so brevity is essential. *Nature's* main formats for original research are Articles and Letters to *Nature*. Lengths of these contributions are specified in terms of numbers of pages they take up. An uninterrupted page of text contains about 1,300 words, so a typical Letter to *Nature*, occupying 2 1/2 pages, can contain about 1,500 words of text and four modest-sized display items (figures and/or tables) with brief legends. There are no formal limits on the number of display items, but extra items must be accompanied by commensurate reductions in the length of the text. Length limits can, however, be varied at the editors' discretion.

Articles are reports of original scientific research whose conclusions represent a substantial advance in understanding of an important problem and are of broad general interest. They do not generally exceed 5 pages of *Nature* and have no more than 50 references. They have a 'heading' of up to 80 words, often rewritten by *Nature's* editors, which advertises the content of the paper in general terms and which should not contain numbers, abbreviations or measurements unless central to the message of the paper. The first three or four paragraphs of the main text should introduce and summarize the findings. Articles have a few subheadings of not more than 45 characters each.

Letters to *Nature* are short reports of original research focused on an outstanding finding with implications of interest to nonspecialists. They should not generally exceed 2 1/2 pages of *Nature*. They have a bold first paragraph of fewer than 180 words summarizing the background, rationale, conclusion and main implication of the work, aimed particularly at nonspecialist readers. Letters should have no more than 30 references.

Preparation of Articles and Letters

All manuscripts should be in English, typed double-spaced on one side of the paper only. Manuscripts (and revised manuscripts) should not be submitted in electronic form.

Manuscripts should be accompanied by two copies of a brief covering letter from the corresponding author, containing a full postal address, telephone and fax numbers, and e-mail address. Articles and Letters contain a statement at the end of the text: "Correspondence and requests for materials to xxxxx", with an e-mail address if desired, followed by the accession numbers or Web site address for supporting datasets. The designated author need not be the senior author, but *Nature* will expect this author to handle matters arising from the published contribution.

Five copies of manuscripts (and revised manuscripts) and figures are required. Each copy of a revised manuscript should be accompanied by a copy of the author's response to referees' and editors' comments on the earlier version. Two copies of relevant papers by any of the authors submitted or in press elsewhere (including papers submitted elsewhere while the *Nature* manuscript is under consideration) should also be enclosed, clearly marked as such; failure to do so may prejudice consideration of the manuscript. Titles should not exceed three lines of 28 characters each (including spaces), and should avoid numerical values, acronyms, abbreviations and punctuation wherever possible.

References should be numbered sequentially as they appear in the text, tables and figure legends. Titles of all cited articles should be provided. Only papers that have been published or submitted to a named journal should be in the reference list: manuscripts in preparation and personal communications should be mentioned in the text with a list of authors. Published conference abstracts may be included in the reference list, but text, footnotes, grant details and acknowledgements may not.

Acknowledgements must be brief and appear after the reference list. *Nature* does not publish grant or contribution numbers.

Tables should each be presented on a separate sheet of paper. Symbols and abbreviations should be defined immediately below the table, fol-

lowed by essential descriptive material, all in double-spaced text.

Figures should not be larger than 22 by 28 cm unless unavoidable, and should not contain more than one panel unless the assembled parts are logically connected. Each copy should be marked with the corresponding author's name and, when known, the manuscript reference number. One photocopy of original figures should be provided. All original figures are returned when *Nature* cannot offer to publish a manuscript, but one copy of the manuscript and the photocopy of the figures are kept in confidential files. Colour figures are welcome, but a contribution towards the cost of reproduction is requested. Inability to pay this charge will not prevent *Nature* from publishing figures for which colour is essential. Once a paper is accepted for publication, electronic versions of figures are preferred.

Methods. Brief details of materials and methods can be given in the text. Otherwise, they should be described at the end of the text in a section entitled "Methods", subdivided by short, bold headings referring to the methods used. This section should not exceed 800 words and ideally should be shorter. In contributions with methods sections, figure legends should not exceed 100 words each, and should be confined to a description of the figure and the symbols used.

Materials. As a condition of publication, authors are required to make materials and methods used freely available to academic researchers for their own use. Supporting datasets must be made available by the authors both directly and by posting on *Nature's* Web site, by deposition in the appropriate public database or by distribution on the Internet.

Submissions. Manuscripts can be submitted to the Editor at *Nature*, Porters South, 4 Crinan St, London N1 9XW or at *Nature*, 968 National Press Building, Washington, DC 20045-1938. Proofs should be returned by express mail to London. Fax numbers and e-mail addresses for all editorial offices are on *Nature's* information page.

Presubmission enquiries for Articles and Letters should be made by e-mail to nature@nature.com, clearly labelled as such. They should include a fully referenced summary paragraph in *Nature* format, and be accompanied by a paragraph stating briefly why the submission will be of interest to a broad general scientific readership. Authors will usually receive a response within two working days of receipt.

Material submitted to *Nature* must not be discussed with the media. Accepted contributions may be discussed with the media a week before the publication date under *Nature's* embargo conditions. (See *Nature* 383, 745; 1996 for further explanation.)

Other contributions

Nature publishes other contributed material as follows:

Correspondence contributions are short comments on topical issues, anecdotal material or reactions of readers to non-technical material published in *Nature*. All are unsolicited; submissions by e-mail to corres@nature.com are preferred (accents should be clearly indicated).

Commentary articles are informal editorials on topical issues of public interest that have some bearing on scientific research. They are often controversial. Unsolicited manuscripts are considered, but proposals can be made to the Commentary Editor by e-mail (nature@nature.com).

News and Views articles inform nonspecialist readers about new scientific advances, as reported either in published papers or at scientific meetings. Most are commissioned, but proposals can be made to the News and Views Editor before the paper is published or the meeting takes place. Detailed guidelines are available on request.

Scientific Correspondence contributions are brief reports of general interest, and are usually peer-reviewed. Priority is given to contributions of fewer than 500 words. Detailed guidelines are available on request. Presubmission enquiries are not considered: contributions for Scientific Correspondence should be submitted by mail, not e-mail.

Review Articles survey recent developments in a topical area of research. They should not generally occupy more than 6 pages of *Nature*.

Progress articles are shorter reviews of particularly topical and fast-moving fields. They do not exceed 4 pages of *Nature*.

Contributors wishing to submit a Review Article or Progress should first send an enquiry or a brief synopsis to the Reviews Coordinator, from whom guidelines are available on request.

Duration of the superwind phase of asymptotic giant branch stars

T. Tanabé*, S. Nishida*†, S. Matsumoto†, T. Onaka†, Y. Nakada‡, T. Soyano‡, T. Ono§, K. Sekiguchi|| & I. S. Glass¶

* Institute of Astronomy, Faculty of Science, The University of Tokyo, Mitaka, Tokyo, 181, Japan

† Department of Astronomy, School of Science, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo, 113, Japan

‡ Kiso Observatory, Institute of Astronomy, Faculty of Science, The University of Tokyo, Mitake-mura, Kiso-gun, Nagano, 397-01, Japan

§ Nishi-Harima Astronomical Observatory, Sayo-cho, Hyogo, 679-53, Japan

|| National Astronomical Observatory, Mitaka, Tokyo, 181, Japan

¶ South African Astronomical Observatory, PO Box 9, Observatory 7935, South Africa

Near the ends of their lives, low- and intermediate-mass stars go through a phase of evolution known as the asymptotic giant branch^{1,2}. This luminous red-giant phase is thought to be terminated by a period of intense mass loss in the form of a superwind³, which leads to the formation of a planetary nebula. Although the effects of mass loss have been studied extensively in many stars, the duration of this phase is not well constrained, because of uncertainties in the distances, masses, ages, and absolute luminosities of the observed stars. On the other hand, the properties of stars in the globular clusters associated with the Magellanic Clouds are not subject to these uncertainties, and so provide an excellent opportunity for studying mass-loss phenomena in a quantitative way. Here we report the discovery of two infrared stars in Magellanic Cloud globular clusters that are undergoing a period of intense mass loss. Those observations, together with those of a previously discovered infrared star, confirm that asymptotic giant branch stars go through a superwind phase, and constrain the duration of this phase to be about 100,000 years.

Stars with masses less than 8–9 times that of the Sun ($1 M_{\odot}$) eventually reach the asymptotic giant branch (AGB). Throughout their evolution along the AGB they lose mass through stellar winds, which increase greatly towards the ends of their lives. During this phase dust grains condense from the ejected gas and a thick circumstellar shell develops around the central star. Stellar light is absorbed by the shell and is re-emitted in the infrared. Thus a mass-losing AGB star is often first detected as an infrared object, implying that infrared observations are necessary for a proper study to be made. More than 80,000 infrared stars were detected in our Galaxy by the IRAS survey⁴ but most of them lack reliable distances and absolute luminosities. In this respect, stars in either of the Magellanic Clouds have the great advantage that their distances are known quite well. The several tens of AGB stars known to exist^{5,6} in globular clusters in the Magellanic Clouds offer the following further advantages for study: (1) all members of a given cluster are formed at the same time, (2) they are chemically homogeneous at birth, and (3) their ages and masses can be estimated. Globular clusters are numerous in the Magellanic Clouds and span a wide range of ages, unlike those in our Galaxy, making them particularly suitable for studying the short-lived later stages of stellar evolution. Most of the cluster AGB stars known up to now are not extremely red, implying that their mass-loss rates are not very high and suggesting that they have still to reach their final evolutionary stage. The lack of heavily mass-losing infrared stars with known mass and luminosity has so far prevented a clear understanding of the transition from the AGB to planetary nebulae.

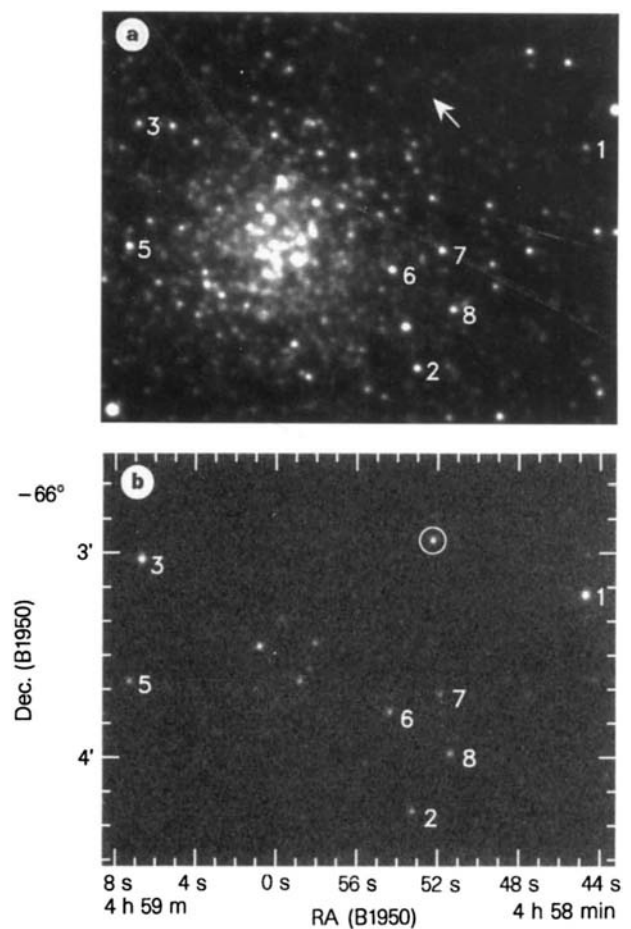


Figure 1 Visual (V-band, 0.55- μm) and infrared (K-band, 2.2- μm) images of NGC1783. The V-band image (a) was taken on the night of 22 December 1992; the K-band image (b) was taken on the night of 6 November 1995. The encircled star in b is NGC1783-IR1. The arrow in a indicates the corresponding position. Several bright AGB stars are also indicated by numbers which correspond to those in ref. 12.

Intermediate-age Magellanic Cloud clusters (SWB⁷ type IV–VI; age, a few $\times 10^8$ to several $\times 10^9$ years) are known to have well developed asymptotic giant branches which in many cases contain carbon stars. Their turnoff mass, the mass of the star at the top of the main sequence and just before leaving it, is estimated⁹ to be (1.4–2.7) $M_{\odot}$. Nine of these clusters, NGC152, NGC416, NGC419, NGC1783, NGC1806, NGC1846, NGC1978, NGC1987 and NGC2121, were chosen to search for infrared AGB stars. The observations reported here were made with the PtSi Astronomical Near-Infrared Camera, (PANIC)^{8–10}, attached to the f/15 Cassegrain focus of the 0.75-m telescope at the Sutherland station of the South African Astronomical Observatory. The detector was a Mitsubishi 1,040 $\times$ 1,040 PtSi near-infrared array¹¹. The image scale was 3.33 pixels per arcsec, giving a total field of view of approximately $5' \times 5'$ per frame. Typical exposure times are 300 s, 600 s and 900 s in the J, H and K bands, respectively. The detection of the infrared stars discussed here has been confirmed by repeated observations. Details of the analysis will be described elsewhere.

Figure 1 shows the visual (V-band, 0.55- μm ; Fig. 1a) and infrared (K-band, 2.2- μm ; Fig. 1b) images of NGC1783. The star we discovered in this cluster (hereafter NGC1783-IR1) is only seen in the infrared and is much brighter in the K band than in the J band (1.2- μm). It is not seen even in the I-band (0.9- μm) images of, for example, Lloyd Evans¹². The infrared star that we found in NGC419 (NGC419-IR1) is even redder than NGC1783-IR1. Our infrared photometric results for both objects are given in Table 1 together

Table 1 Photometric summary of infrared stars

Objects	Coordinate (1950)		<i>J</i> (mag)	<i>H</i> (mag)	<i>K</i> (mag)	<i>M</i> _{bol} (mag)	Date of observation
	RA (h min s)	Dec (° ' ")					
NGC1783-IR1	4 58 52.2	− 66 2 56	14.0	12.6	10.8	− 4.4	6–7 Nov. 1995
NGC419-IR1	1 6 42.8	− 73 8 44	14.7	12.6	10.9	− 4.9	27 Jun. 1995
NGC1978-IR1*	5 28 37.0	− 66 16 11	13.4	11.8	10.3	− 4.9	2–4 Jan. 1996

The data reduction was carried out with the Image Reduction and Analysis Facility (IRAF) and the photometry was performed with DoPHOT¹⁵. Uncertainties in *J*, *H* and *K* are ± 0.2 mag; uncertainties in *M*_{bol} are ± 0.2 mag. The positions were determined in two steps. First, the positions of the several bright AGB stars in each cluster were measured on UK Schmidt Southern Infrared Atlas (I Plate, no. 29 for NGC419 and no. 85 for NGC1783 and NGC1978, respectively) with the XY measuring machine at Kiso Observatory. Second, infrared stars were measured on our infrared images referring to these stars. The positional accuracy is estimated to be within 1 arcsec. The bolometric magnitudes (*M*_{bol}) were obtained by fitting the blackbody curve to the *J*, *H* and *K* values. The distance moduli were taken as 18.5 for the Large Magellanic Cloud and 18.9 for the Small Magellanic Cloud. These values may not be definitive because of possible infrared excess over the blackbody curve at longer wavelengths due to dust grains and because of the variability mentioned in the text.

*NGC1978-IR1 is identical to no. 101 in NGC1978 observed by Ferraro *et al.*¹⁹.

with their estimated bolometric magnitudes. In Fig. 2, we have plotted the energy distributions of the infrared stars. These newly discovered infrared stars have similar colours to NGC1978-IR1 (third object in Table 1) mentioned by Frogel *et al.*⁵. We have measured their precise positions, which we present in Table 1. Combined with the fact that all the three clusters contain carbon stars, the extreme redness of the infrared stars suggests that they are carbon-rich Mira variables¹³. In this respect it should be noted that the bolometric magnitudes of these infrared stars are near to, in the case of NGC1783-IR1, or higher than, in the cases of NGC419-IR1 and NGC1978-IR1, the transition luminosity from M-type to C-type stars⁵. Our preliminary analysis also suggests that they are variable in the infrared.

We therefore conclude that these infrared stars are luminous AGB stars which clearly stand apart from the other, visible, AGB stars in that they have much thicker dust shells resulting from heavier mass-loss. We thus believe that they are in the final stage of their AGB evolution, the superwind phase. These infrared stars, together with other AGB stars, form a sequence leading towards planetary nebulae and constitute direct evidence that an AGB star goes through a superwind phase.

According to the scheme of Renzini and Buzzoni¹⁴, the lifetime of this phase can be determined statistically by the contribution of the luminosity from the mass-losing stars to the total luminosity of the observed globular clusters.

From equation (14) of ref. 14 ($n_i = 1.8 \times 10^{-11} L_T t_i$, where n_i is the number of stars in any given post-main-sequence stage, L_T is the

total bolometric luminosity surveyed and t_i is the duration of that stage), we derive 1×10^5 years as the duration of this superwind stage. (Here we estimated that L_T is 1.2×10^6 times the solar luminosity, by summing up the individual cluster's luminosity calculated from the apparent bolometric magnitudes³ with the distance moduli given in Table 1.) This duration implies that the mass-loss rate of these infrared stars is $\sim 1 \times 10^{-5} M_\odot \text{ yr}^{-1}$ because we are observing the evolution of stars whose initial masses are 1.4–2.7 $M_\odot$. This value of mass-loss rate is somewhat smaller than estimates based on observations of planetary nebulae³ and on heavily obscured Galactic carbon stars^{15,16} (several to ten times $10^{-5} M_\odot \text{ yr}^{-1}$). It should be remembered that these other estimates include uncertainties in the distances of the objects. Our mass-loss rate is more consistent with recent stellar evolution calculations¹⁷ with realistic mass-loss which given the duration of the superwind phase as 5×10^4 to 1.5×10^5 years for the case of $M \sim 2 M_\odot$.

Further examination of the infrared stars that we have found will help our understanding of the late stages of evolution of low- to intermediate-mass stars. Observations with the Infrared Space Observatory (ISO) should provide the full spectral energy distribution of these objects up to 20 μm , which will enable us to estimate their luminosities more accurately. Further survey work in the infrared should lead to a much more precise determination of the duration of the mass-loss phase. These quantities can be used directly to test the current theory of stellar evolution. Our observations also suggest the possibility that there may be a number of highly luminous objects still hidden behind thick dust shells. $\square$

Received 5 July; accepted 12 December 1996.

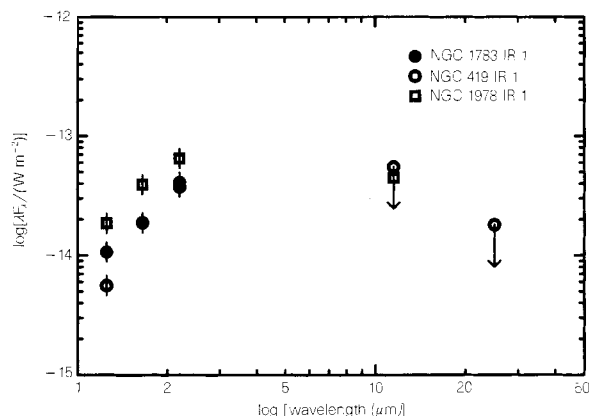


Figure 2 Spectral energy distributions of infrared stars. NGC1783-IR1, filled circle; NGC419-IR1, empty circle; NGC1978-IR1, empty square. High-quality IRAS data of IRAS F01067–7308 which was detected towards NGC419 and of IRAS F05287–6616 towards NGC1978 are also plotted. These values should be regarded as upper limits because IRAS did not resolve these objects and they may contain contributions from other stars. No IRAS source was detected towards NGC1783.

1. Iben, I. Jr & Renzini, A. *Annu. Rev. Astron. Astrophys.* **21**, 271–342 (1983).
2. Habing, H. J. *Astron. Astrophys. Rev.* **7**, 97–207 (1996).
3. Renzini, A. in *Physical Processes in Red Giants* (eds Iben, I. Jr & Renzini, A.) 431–446 (Reidel, Dordrecht, 1981).
4. Chester, T. in *Light on Dark Matter* (ed. Israel, F. P.) 3–22 (Reidel, Dordrecht, 1986).
5. Frogel, J. A., Mould, J. & Blanco, V. M. *Astrophys. J.* **352**, 96–122 (1990).
6. Westerlund, B. E., Azzopardi, M., Breysacher, J. & Rebeiro, E. *Astron. Astrophys. Suppl. Ser.* **91**, 425–451 (1991).
7. Searle, L., Wilkinson, A. & Bagnuolo, W. G. *Astrophys. J.* **239**, 803–814 (1980).
8. Glass, I. S., Carter, D. B., Sekiguchi, K. & Nakada, Y. in *Infrared Astronomy with Arrays: The Next Generation* (ed. McLean, I. S.) 285–286 (Kluwer, Dordrecht, 1994).
9. Glass, I. S., Sekiguchi, K. & Nakada, Y. in *New Developments in Array Technology and Applications* (eds Davies, Phillip, A. G., Jones, K. A. & Upgren, A. R.) 109–116 (Kluwer, Dordrecht, 1995).
10. Tanabé, T. *et al. Proc. SPIE* **2744**, 110–114 (1996).
11. Yutani, N. *Tech. Dig. of IEDM* 175–178 (1991).
12. Lloyd Evans, T. *Mon. Not. R. Astron. Soc.* **193**, 87–96 (1980).
13. Feast, M. W. *et al. Mon. Not. R. Astron. Soc.* **201**, 439–450 (1982).
14. Renzini, A. & Buzzoni, A. in *Spectral Evolution of Galaxies* (eds Chiosi, C. & Renzini, A.) 195–235 (Reidel, Dordrecht, 1986).
15. Knapp, G. R. & Morris, M. *Astrophys. J.* **292**, 640–669 (1985).
16. Knapp, G. R. *Astrophys. J.* **293**, 273–280 (1985).
17. Vassiliadis, E. & Wood, P. R. *Astrophys. J.* **413**, 641–657 (1993).
18. Schechter, P. L., Mateo, M. & Saha, A. *Publ. Astron. Soc. Pacif.* **105**, 1342–1353 (1993).
19. Ferraro, F. R. *et al. Mon. Not. R. Astron. Soc.* **272**, 391–422 (1995).

Acknowledgements. We thank M. Ueno (University of Tokyo) and M. Kimata at the Mitsubishi Electric Corporation, Advanced Technology R & D Center for information about the PISi chip; D. B. Carter at SAO for his work on the electronics of PANIC; and other members of the SAO technical staff for their help. We also thank T. Lloyd Evans for a critical reading of the manuscript.

Correspondence should be addressed to T.T. (e-mail: ttanabe@mtk.iao.s.u-tokyo.ac.jp).

Dendritic growth of electrohydrodynamic convection in a nematic liquid crystal

J. T. Gleeson

Department of Physics and Astronomy, The University of Calgary, Calgary, Alberta T2N 1N4, Canada

When a liquid is supercooled and begins to crystallize, the stable solid phase grows and penetrates into the metastable liquid phase. The resulting crystallites often appear as finger-like dendrites^{1,2} which branch as they grow into complex structures, similar in appearance to the arms of a snowflake³. The basic pattern of dendritic growth is common to many systems undergoing phase transitions, is observed during viscous fingering and electrochemical deposition, and plays an important role in determining the strength of cast metals⁴. Here we report the results of experiments which show that dendritic growth can also occur in the very different context of electrohydrodynamic convection. In a nematic liquid crystal, convective flow can be induced by the application of a sufficiently strong electric field. We find that, at the onset of convection, the convective state can invade the equilibrium state in the form of dendritic patterns. These results, which cannot be explained in terms of the existing theory for electrohydrodynamic convection, imply that the phenomenon of dendritic growth is far more ubiquitous than was previously suspected.

Dendrites are a common morphology for diffusion-controlled crystal growth in the presence of anisotropy^{5–9}. The most common example is the nucleation and growth of a crystal in a supercooled melt. For this crystal to grow, some quantity that is generated at the interface between crystal and melt during freezing, for example the latent heat of crystallization, must diffuse away from that interface. Like fins on a thermal radiator, the sharp tip of the dendrite promotes diffusion of heat away from itself, allowing the crystal to grow rapidly in the direction in which the tip is pointing. One characteristic of dendrites is that the shape of the tip does not change in the frame of reference moving at the tip speed. As one of the most fundamental questions about dendritic growth is what determines the shape of the crystal–melt interface, the properties of that interface are significant. Two properties have been found to be particularly important: surface tension and anisotropy. In an isotropic environment, if a non-zero surface tension is included, no shape-preserving, steady-state dendrite solution to the diffusion problem exists; only the presence of anisotropy makes dendrite solutions possible. Experiments on Saffman–Taylor fingering in Hele–Shaw cells (in which a viscous fluid confined between parallel plates is displaced by injecting a less-viscous fluid) using both externally imposed anisotropy¹⁰ and anisotropic liquids¹¹, confirmed that a sufficient degree of anisotropy must be present for steady-state dendrites to exist.

Nematic liquid crystals are anisotropic fluids in which the constituent molecules have orientational but no positional order. Electrohydrodynamic convection (EHC) in nematics is a popular system^{12,13} in which to investigate pattern formation out of equilibrium¹⁴. EHC is in many ways similar to a better known system: thermally induced Rayleigh–Bénard convection. Because a nematic is intrinsically anisotropic, a separation of charge can be induced in the direction perpendicular to that in which an electric field is applied. The force on the separated charge in the electric field causes convective flow when the strength of the applied electric potential exceeds some critical value.

Results^{15,16} on EHC in which a magnetic field is applied parallel to

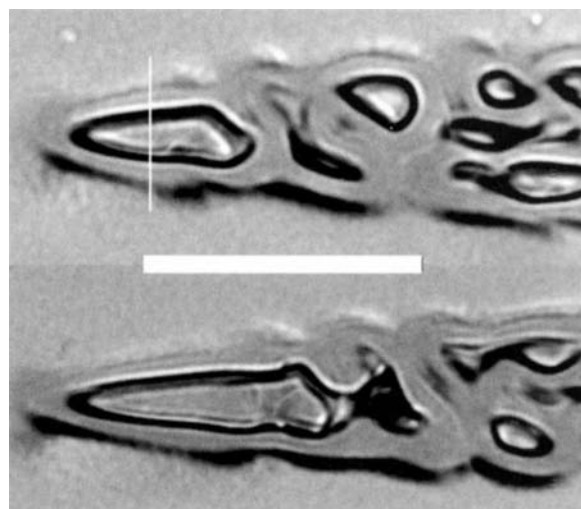


Figure 1 Two shadowgraph images of single dendrites of convection growing into the Freedericksz distorted state. This structure emerges spontaneously shortly after the applied voltage, V , is quickly changed from zero to some value above the critical voltage, V_A , above which convection can be observed. The experimental configuration is the 'classic' setup^{12,29}, in which the liquid crystal is confined in a parallel-plate geometry with the director (at zero applied fields) constrained to lie in one direction, which we define as x . In addition we apply a constant magnetic field parallel to the applied electric field. The liquid crystal sample is methoxy-benzylidene butylaniline (MBBA) and the plate separation is $50\text{ }\mu\text{m}$. This image is the view along the perpendicular to the plates. The white bar represents $200\text{ }\mu\text{m}$; the thin white line is a reference line for Fig. 2. The lateral dimensions of the conducting area are $2\text{ mm} \times 2\text{ mm}$. The frequency of the applied electric field is 50 Hz . The standard material parameters³⁰ for MBBA at $25\text{ }^\circ\text{C}$ are assumed to hold; the electrical conductivities parallel and perpendicular to the director, measured *in situ*, are $11.7 \times 10^{-8}\text{ }\Omega^{-1}\text{ m}^{-1}$ and $8.5 \times 10^{-8}\text{ }\Omega^{-1}\text{ m}^{-1}$ respectively. For the upper image the applied r.m.s. voltage is 10.544 V , and for the lower it is 10.241 V . The magnetic field is 3.544 kG .

the applied electric field (both along z) have revealed that the transition from the no-convection, or quiescent, state to the convection state is subcritical (that is, discontinuous; as the applied voltage is slowly raised, the amplitude of convection jumps abruptly at a critical voltage¹⁷) at magnetic fields greater than about $1.5 H_F$, where H_F is the critical field for the Freedericksz transition (a reorienting of the nematic director under the influence of an external electric or magnetic field; see ref. 16). Moreover, the transition becomes ever more strongly subcritical as H is increased above this value. Thus, at low or zero magnetic field, the transition to convection is analogous to a second-order phase transition, whereas at high magnetic fields, as H increases, the transition to convection behaves like a phase transition becoming more strongly first order.

The most important difference between the transitions to convection at low and high magnetic fields is unquestionably the nature of the quiescent state. At low magnetic fields this state is planar-aligned, in which the nematic director, a unit vector field describing the direction of average molecular orientation, is everywhere parallel to a common direction, x , and perpendicular to the applied electric field. At high magnetic fields, the director in the quiescent state remains parallel to x only near the glass substrates. Midway between these substrates the director tilts towards z ; this tilt increases with $H - H_F$. Where the director is parallel to the applied fields, the system is isotropic in the plane perpendicular to the fields. In the experiments presented here, the director is parallel to the applied fields in the region midway between the glass substrates, and perpendicular to the applied fields close to the substrates. The director rotates between these two directions over a distance called

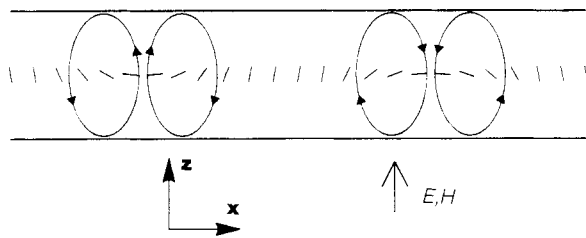


Figure 2 Proposed configuration for the dendritic structure. This represents a cross-section through the sample cell along the thin white line shown in Fig. 1. The direction of fluid flow and the nematic director at the midway point between the electrodes are shown, as are the externally applied fields.

the magnetic coherence length¹⁸, which is proportional to $1/H$. Thus, in these experiments, the coherence length, which determines the size of that region of the sample in which the system is anisotropic and hence the effective anisotropy, is continuously variable as H varies. For all the experiments reported here, the magnetic coherence length was less than a tenth of the distance between the substrates, so the system was anisotropic within a region making up less than 20% of the volume in which convection occurred.

In Fig. 1 we see a still, shadowgraph image of a dendrite of the convection state growing into the quiescent state. In a shadowgraph, variation in refractive index (here caused by the director distortion) causes focusing and defocusing of parallel light rays, allowing convective flow to be visualized¹⁹. (For a motion picture of the dendritic growth, see ref. 20). Figure 2 shows a proposed configuration of the flow field and the nematic director for this dendrite. If the applied voltage, V , is not too large the dendrites appear in the same location every time, probably because of small inhomogeneities in the sample cell. At greater V , however, the dendrites grow very rapidly in many places simultaneously. As we are interested in examining single dendrites, this represents a practical upper limit on the size of V . Furthermore, although Fig. 1 shows only part of a single dendrite, convection dendrites always originate in pairs which then travel in opposite directions at the same speed. This direction is oriented at about 80° to x . Like dendrites in other systems, these dendrites show secondary instabilities. In the frame moving with the dendrite tip, these instabilities occur in a region located some distance from the tip. The distance decreases as the tip speed decreases. These instabilities of the dendrite are analogous to side-branches in crystalline dendrites.

The relationship between the speed of a dendrite tip, the shape of the tip and the distance from equilibrium (the dimensionless undercooling, Δ) has been the subject of much investigation. In the absence of surface tension between the dendrite and the host into which it grows, the boundary value problem for the diffusion equation may be solved exactly (the Ivantsov solution)^{21,22}; however, in this limit, the tip speed and tip curvature are not independent. Specifically, one finds that only the Péclet number, $P \equiv \rho v/2D$, where ρ is the radius of curvature at the dendrite tip, v is the tip speed and D is the relevant diffusion constant, is determined by Δ , so that neither v nor ρ is uniquely determined. In two dimensions, the Ivantsov solution yields dendrites (ignoring any side-branches) that are parabolas; in the absence of surface tension, however, these parabolic dendrite solutions are linearly unstable. When finite surface tension, but no anisotropy, is incorporated into the boundary conditions for the diffusion equation, no steady-state solution can be found. In this situation, both experiments and numerical simulations find that dendrite tips are not stable, and continuously split in two^{10,23}. Only when sufficient anisotropy is also present does a stable dendrite solution exist; this dendrite can only grow in a direction specified by the direction of anisotropy.

In Fig. 3 we plot the tip speed of a large number of dendrites against the square of the voltage (V) applied to the sample cell. For

an individual dendrite, the uncertainty in determining the speed is always less than 1%; in the figure, at a given voltage, we see variations in tip speeds of 30% or more, especially at larger speeds. Thus, the tip speed is not uniquely selected by V , in contrast to the results of careful measurements on crystalline dendrites²⁴. The dendrite tips are well described by parabolas, although accurately measuring the tip radius of curvature is difficult. Figure 1 represents the sharpest image of the dendrite that can be obtained. The intrinsic width of the dark region is comparable to the tip radius, so any attempt to fit the shape of this region to a specific form is inherently suspect. Nonetheless, although difficult to measure, ρ remains roughly constant at about $3 \pm 0.7 \mu\text{m}$, except at the highest observed speeds.

One difficulty in describing this system within the context of either crystal growth or Saffman–Taylor dendrites is that analogies to quantities such as surface tension, dimensionless undercooling and so on are not well described. Straightforwardly fitting to the Ivantsov solution, in which we define the dimensionless undercooling as $(V^2 - V_0^2)/V_0^2$, yields $V_0 = 7.44 V_{\text{ms}}$, $V_A = 8.33 V_{\text{ms}}$. Assuming that ρ is constant, the coefficient of v is $\rho/2D$. Using our rough value of ρ , we find the relevant diffusion constant to be $3.4 \times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$; that is within 50% of the orientational diffusion constant (the ratio of the splay elastic constant to the rotational viscosity¹⁸). Thus, the analogy with the standard dendrite problem seems to be complete. The diffusion field controlling the dendrite growth is the orientation of the nematic director. Also shown in Fig. 3 is the fit to the Ivantsov solution under the assumption that ρ is constant. Despite the known shortcomings of the Ivantsov solution, this relationship appears to describe the data well over its entire range.

These observations indicate that dendritic growth is not confined to either phase transitions or Saffman–Taylor fingering, but is a more pervasive growth mechanism than previously thought. Indeed, all known systems in which dendrites are observed can be cast in terms of a Stefan problem: that is, a boundary value problem in which the position of the boundary at some point in the future is determined by the solution to the problem at earlier times. Currently, we are not aware of any theory of electrohydrodynamic convection, or any other variety of convection, that can be cast in these terms. Also, the outline of these dendrites can be thought of as the transition region between the area in which the liquid crystal is stationary and the area in which convective flow occurs. Such transition regions ('fronts') are well known in non-equilibrium pattern formation. The propagation of fronts has been much studied^{25–27}, but the transition region between convection and no-convection in this case differs from familiar cases of front propagation because it is controlled by diffusion. We may conclude this

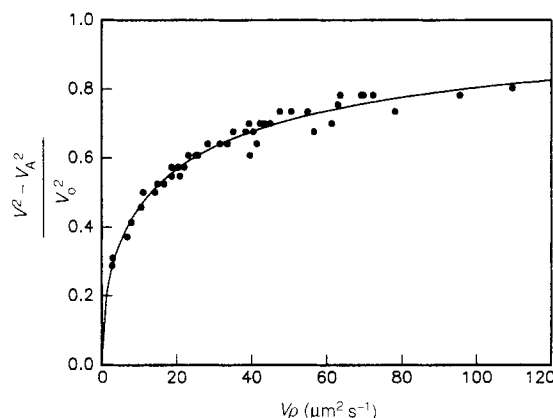


Figure 3 Square of applied voltage as a function of the product of tip speed v and a tip radius of curvature ρ for a large number of dendrites. The solid line is a fit to the relationship between the dimensionless undercooling Δ and the Péclet number P yielded by the two-dimensional Ivantsov solution: $\Delta = \sqrt{\pi P} \exp(P) \text{erfc}(\sqrt{P})$. We have treated the tip radius of curvature as a constant with value $3 \mu\text{m}$.

because the dendrites of convection reported here contain all the elements of the crystalline dendrite problem, and the dynamics of crystalline dendrites are controlled diffusion. To our knowledge, front propagation of this type has not been previously observed.

We can conjecture what ingredients are necessary for similar dendrites to be observed in other popular systems for studying pattern formation during convection. Two features seem to be particularly important. The transition to convection must be subcritical, and there must be underlying anisotropy. Thermal convection driven by surface tension²⁸ can show a subcritical transition to convection; if underlying anisotropy is imposed on this system, might that be sufficient to produce dendritic growth?

One of the most important remaining questions about the dendrites reported here is whether they show a morphology transition as either the anisotropy or the distance from equilibrium is decreased. The present experiments show no evidence of such a transition. □

Received 6 March; accepted 23 December 1996.

- Huang, S. C. & Glicksman, M. E. *Acta Metall.* **29**, 717–734 (1981).
- Langer, J. S. *Rev. Mod. Phys.* **52**, 1–28 (1980).
- Kepler, J. *De Nive Sexangula* (1611); transl. Hardie, C. (Clarendon, Oxford, 1966).
- Kurz, W. & Fisher, D. F. *Fundamentals of Solidification* (Trans. Tech. Aedermannsdorf, 1966).
- Kessler, D. A., Koplik, J. & Levine, H. *Adv. Phys.* **37**, 255–339 (1988).
- Langer, J. S. *Science* **243**, 1150–1156 (1989).
- Ben-Jacob, E. & Garik, P. *Nature* **343**, 523–530 (1990).
- Brener, E. A. & Mel'nikov, V. I. *Adv. Phys.* **40**, 53–97 (1991).
- Ben-Jacob, E. *Contemp. Phys.* **34**, 247–273 (1993).
- Ben-Jacob, E. *et al. Phys. Rev. Lett.* **55**, 1315–1318 (1985).
- Buka, A., Palfy-Muhoray, P. & Rac, Z. *Phys. Rev. A* **36**, 3984–3989 (1987).
- Rehberg, I., Winkler, B. L., de la Torre Juarez, M., Rasenat, S. & Schöpf, W. *Adv. Solid State Phys.* **29**, 35–52 (1989).
- Kramer, L. & Pesch, W. in *Pattern Formation in Liquid Crystals* (eds Buka, A. & Kramer, L.) (Springer, New York, 1996).
- Cross, M. C. & Hohenberg, P. C. *Rev. Mod. Phys.* **65**, 851–1112 (1993).
- Gleeson, J. T. & Todorović-Marinić, D. *Molec. Cryst. Liq. Cryst.* **256**, 327–335 (1995).
- Hertrich, A., Pesch, W. & Gleeson, J. T. *Europhys. Lett.* **34**, 417–422 (1996).
- Gleeson, J. T. *Phys. Rev. E* **4**, 6424–6429 (1996).
- deGennes, P. G. & Prost, J. *The Physics of Liquid Crystals*, 2nd edn (Clarendon, Oxford, 1993).
- Rasenat, S., Hartung, G., Winkler, B. L. & Rehberg, I. *Exp. Fluids* **7**, 412–420 (1989).
- <http://www.phas.ucalgary.ca/~gleeson/dendrite.html>
- Ivantsov, G. P. *Dokl. Akad. Nauk. SSSR* **58**, 567–569 (1947).
- Brener, E. A. & Mel'nikov, V. I. *J. Phys. France* **51**, 157–166 (1990).
- Kessler, D. A., Koplik, J. & Levine, H. *Phys. Rev. A* **30**, 2820–2823 (1984).
- Glicksman, M. E., Schaefer, R. J. & Ayers, J. D. *Metall. Trans. A* **7**, 1747–1759 (1976).
- Dee, G. & Langer, J. S. *Phys. Rev. Lett.* **50**, 383–386 (1983).
- van Saarloos, W. *Phys. Rev. A* **37**, 211–229 (1988).
- van Saarloos, W. *Phys. Rev. A* **39**, 6367–6390 (1989).
- Schatz, M. F., VanHook, S. J., McCormick, W. D., Swift, J. B. & Swinney, H. L. *Phys. Rev. Lett.* **75**, 1938–1941 (1995).
- Joets, A. & Ribotta, R. *Phys. Rev. Lett.* **60**, 2164–2167 (1988).
- Bodenschatz, E., Zimmermann, W. & Kramer, L. *J. Phys. France* **49**, 1875–1899 (1988).

Acknowledgements. We thank W. van Saarloos, R. B. Hicks, D. J. Knudsen, J. L. Bechhoefer, M. Lamoureux, A. R. Taylor, P. Palfy-Muhoray and E. Ben-Jacob for discussions. This work was supported by NSERC and the University of Calgary.

Correspondence should be addressed to J.T.G. (e-mail: gleeson@phas.ucalgary.ca)

Hydrothermal growth of diamond in metal–C–H₂O systems

Xing-Zhong Zhao*, Rustum Roy, Kuruvilla A. Cherian & Andrzej Badzian

Intercollege Materials Research Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

The first report of synthetic diamond involved a high-pressure high-temperature (HPHT) process in which diamond was the thermodynamically stable phase¹. Subsequent attempts to make diamond at less extreme conditions culminated in the rapid

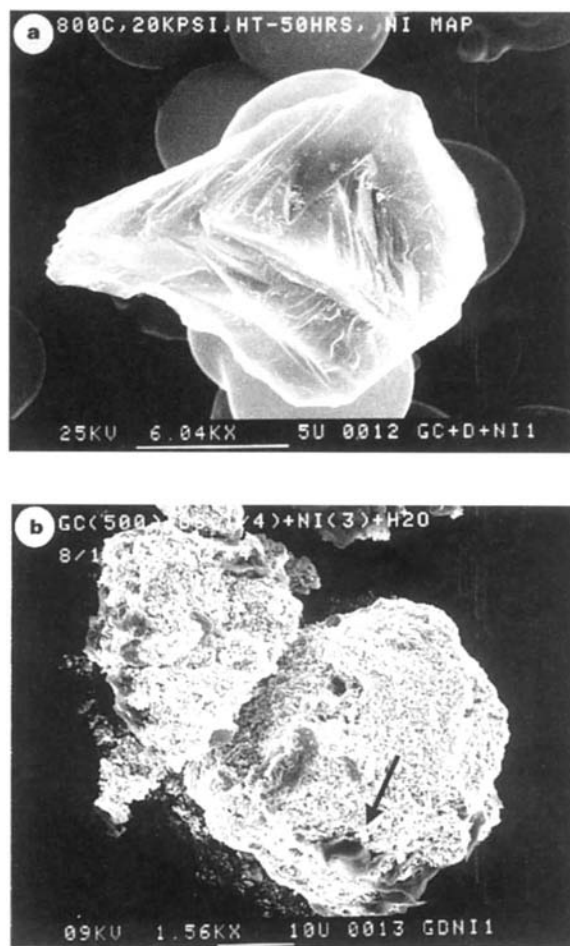


Figure 1 Typical morphology of the product of the hydrothermal synthesis. **a**, Particle, several micrometres in size. **b**, Large aggregate, typically tens of micrometres in size but can be >100 µm. Scale bars: **a**, 5 µm; **b**, 10 µm.

growth of diamond films by chemical vapour deposition^{2,3}. But the question of whether diamond might be grown in hydrothermal conditions mimicking those under which it is formed in the Earth has been long debated^{4–6}. It seems reasonable to suppose that metals might play a catalytic or solubilizing role in this context, given their role in the HPHT method¹, in a recent low-pressure solid-state synthetic approach⁷ and in the recrystallization of diamond in the Ni–NaOH–C system⁸. Hydrothermal synthesis of diamond has been explored at some length^{9–13}, but with equivocal results. Here we report evidence from spectroscopic, diffraction and microscopic techniques which suggest that aggregates, tens of micrometres in size, of small diamond crystals can be grown in a hydrothermal environment from a mixture of carbon, water and metal (usually pure nickel). Crucially, we have been able to distinguish new diamond from the diamond seeds added to nucleate new growth.

Szmanski *et al.*¹³ have claimed hydrothermal growth of diamond, but no details were given on the composition of the liquid and the characterization of the phases. The synthesis of diamond via the SiC+H₂O reaction has also been claimed by Gogotsi *et al.*¹⁴ but the existence of a ternary phase in the Si–O–C system with a Raman peak around 1,330 cm^{–1} (ref. 11) allows for an alternative explanation of the results. Roy *et al.*¹² have made a detailed study of the hydrothermal growth of diamond in the C–H–O and C–H–O–halogen systems, showing considerable X-ray diffraction (XRD) evidence for diamond growth. But con-

* Permanent address: Department of Materials, Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China.

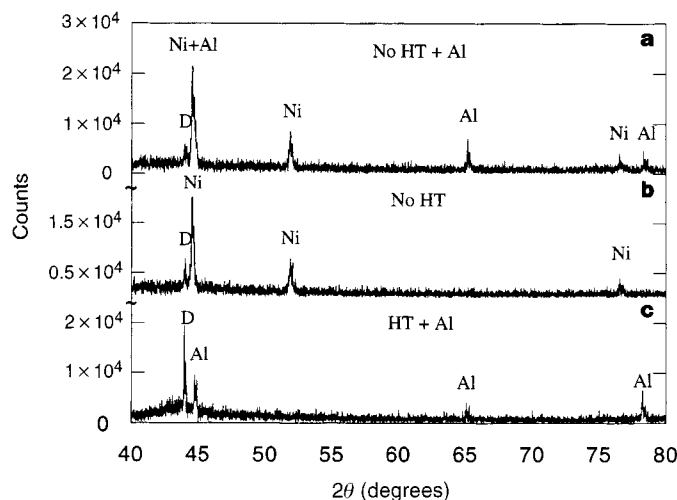


Figure 2 X-ray diffraction patterns of **a**, sample after hydrothermal (HT) run, 10 mg in weight with 0.5 mg Al powder added as internal standard; **b**, sample of 10 mg glassy carbon mixed with 0.2 mg 0.25- μm diamond seed and 0.3 mg Ni powder (that is, the same as actually used in the run) before the hydrothermal run, and **c**, sample after 0.5 mg Al was added to the latter mixture. D indicates main diamond peak. The data clearly support the appearance of new diamond after the hydrothermal run.

sistent evidence from several different characterization tools has not previously been accumulated.

Our present experiments were done with conventional Roy-Tuttle bombs in the hydrothermal pressure-temperature regime near 800 °C and 1.4 kbar for a typical run. We mixed 3 wt% powdered nickel (99.7% pure, $\sim 3\ \mu\text{m}$, flake-like, Aldrich) with 95 wt% glassy carbon ($\sim 3\ \mu\text{m}$, spherical, prepared by S. Yamada, Nishi-Tokyo University, Japan), and 2 wt% 0.25- μm diamond seeds (Warren Diamond Powder Co., Olyphant, PA, USA) and water (50–100 wt% of the glassy carbon), and sealed the mixture into gold tubes. The size of the seeds was checked by scanning electron microscopy (SEM): although there was some distribution around the average of 0.25 μm , they were all smaller than 1 μm . After the run, each sample was analysed using a Scintag DMC-105 X-ray diffractometer (XRD), a ISI-DS 130 scanning electron microscope (SEM) and a ISA U-1000 micro-focus Raman spectrometer with green argon laser excitation and a 3- μm diagnostic spot.

In typical runs, the original black powder charge turns grey and large bonded aggregates are observed. The morphology of recovered powders after 50 hours of hydrothermal reaction was irregular with some anhedral particles. Figure 1a shows the typical morphology of these particles. The Ni-map and energy-dispersive X-ray (EDX) element analysis over a wide range did not show any evidence for nickel or other metals above background in these particles, thereby ruling out the possibilities that the particles shown in Fig. 1a were metal oxide or metal carbide. Individual particles grow larger after longer (100-hour) hydrothermal runs, but show predominantly a different morphology (Fig. 1b). We have not observed these latter aggregates in some 500 hydrothermal runs without the metal. Although their surface appears rough, one can also identify relatively smooth or flat regions (arrow in Fig. 1b). No metallic elements were detected by crude EDX analysis in these larger aggregates. As only glassy carbon and 0.25- μm diamond seeds were the other starting materials, these results demonstrate that something completely different from the starting materials had been grown during the hydrothermal process.

The intensity of the diamond peak in XRD patterns increases substantially after the hydrothermal runs (Fig. 2). No nickel or

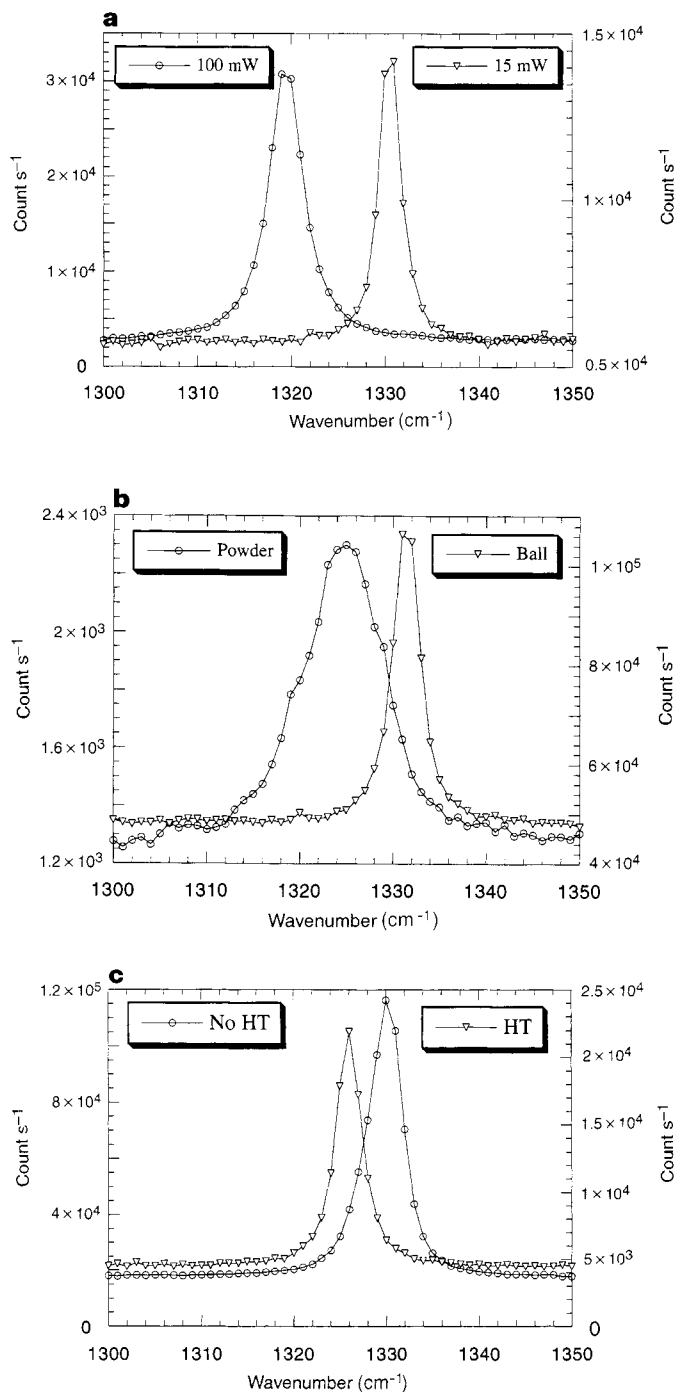


Figure 3 Raman spectra. **a**, Effect of laser power, recorded on a particle $\sim 7\ \mu\text{m}$ in size similar to that shown in Fig. 1a. **b**, Spectra of a large aggregate ('ball') and pressed powders. **c**, Effect of hydrothermal treatment on spectrum of particles.

nickel oxide was observed by XRD, although some nickel-containing materials are observed in the smaller aggregates by EDX analysis and some metal oxide was observed by XRD when other metals were used instead of Ni. Because the XRD patterns are taken from the same volume of material with the same ratio of glassy carbon to 0.25- μm diamond seeds in the starting materials, the peak height of diamond is a reasonable semi-quantitative measure of the concentration of diamond in the initial and final samples. These data therefore indicate that a substantial amount of new diamond has grown during the hydrothermal process.

Raman spectra were obtained from samples after hydrothermal

treatment and those with pure diamond seeds. It was possible to focus the probe laser beam on the individual particles or on aggregates. On individual particles the Raman peak was found to be shifted gradually from 1,331 to 1,319 cm^{-1} as the laser power was increased from 15 to 100 mW using a circular variable beamsplitter and maintaining the laser beam and sample position unchanged (Fig. 3a). The same Raman shift with increasing power was observed repeatedly on more than a half dozen different particles with similar size ($<10\text{ }\mu\text{m}$) in the hydrothermal samples, and we have found¹⁵ that the Raman peak of diamond is affected by both laser power and particle size due to thermal effects. The Raman spectra from the large aggregates, meanwhile, consistently showed only the sharp 1,331 cm^{-1} diamond peak even at laser powers of 100 mW (Fig. 3b). But densely pressed compactions of pure 0.25- μm diamond seeds do show the downshift of the Raman peak to around 1,319 cm^{-1} at a laser power of 100 mW (ref. 15), because the thermal contact between grains is sufficiently poor to produce the downshift. This implies that there is enough new diamond growth and/or new diamond bonding to give good thermal contact between grains in the larger hydrothermally grown aggregates.

To exclude a possible role of impurities present in the seeds, Raman spectra of 5- μm pure diamond seeds and particles larger than 5 μm in the hydrothermal specimen were recorded with a laser power of 50 mW (Fig. 3c). A peak shift of 6 cm^{-1} occurred for the hydrothermal particles, but of only 2 cm^{-1} for the 5- μm pure diamond seeds at the same laser power. Raman spectra obtained from pure pressed diamond seeds of different sizes show that peak shift for the pure seeds larger than 5 μm is $<2\text{ cm}^{-1}$ with laser power no larger than 100 mW. These results also indicate that particles several micrometres in size observed by SEM are different from the original diamond seeds, although they have a vaguely similar morphology.

Figure 4a shows another line of evidence: the Raman and luminescence spectra. These were recorded on the same 0.25- μm pure diamond seeds used throughout, and on a large aggregate similar to that in Fig. 1b, at a power of 30 mW. At least three major differences appear between these spectra. First, there is a huge luminescence peak recorded from the aggregate in the hydrothermal samples, but only a very weak luminescence from the pure 0.25- μm diamond seeds. The luminescence spectrum of the aggregate is very similar to that of diamond films produced by chemical vapour deposition¹⁶, CVD (Fig. 4b). In CVD work it is assumed¹⁷ that crystal defects or impurities are responsible for the features of the Raman and luminescence spectra. This result confirmed again that some kind of new diamond had grown in the hydrothermal process. Second, there is an intensity difference of at least one order of magnitude in the characteristic diamond peaks between the pure 0.25- μm diamonds seeds and the large aggregates. Raman data obtained on pure diamond seeds with different sizes has shown that the larger the size of pure diamond particles, the stronger the intensity of characteristic diamond peaks. Hence the strong diamond peak of the aggregates suggests that these must have either a dense core or well developed bonding, despite the aggregate-like surface morphology. Third, peak shifts in the pure diamond seeds is more than three times larger than that of aggregate in the hydrothermal sample. This is also attributed to the 'effective' size difference between the pure seeds and the aggregate. These three differences in Raman data strongly support the conclusion that hydrothermally grown diamond is substantially different from the seeds.

Thus these lines of evidence imply that the hydrothermal method produces diamond particles several micrometres in size, or aggregates of diamond several tens of micrometres in size, which are clearly different from the 0.25- μm diamond seeds. Although we have not succeeded in growing such particles in the absence of seeds, we can distinguish the seeds from the hydrothermally grown diamond. The key aspect in this work is the use of a metal (primarily nickel, but similar results have been obtained by using platinum and

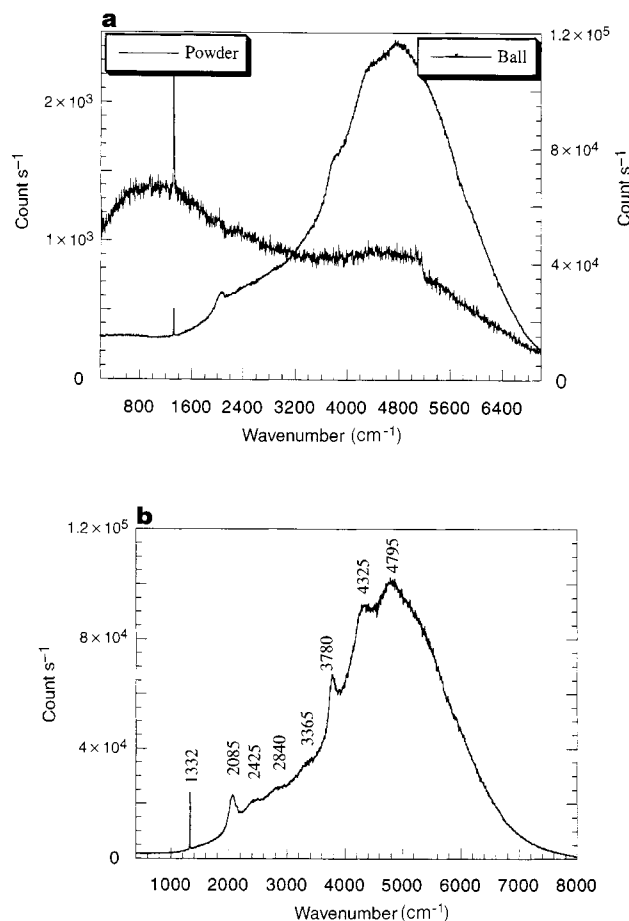


Figure 4 Raman and luminescence spectra **a**, powders and a hydrothermally produced large aggregate (ball), and **b**, CVD diamond film. Note the luminescence typical of synthetic diamond.

iron). At this stage we do not know the mechanism for the hydrothermal growth process; diamond precipitating from a Metal- C_xH_y metallic liquid, or the presence of atomic hydrogen from the reaction $\text{Ni} + \text{H}_2\text{O} = \text{NiO} + \text{H}_2$ might be involved. $\square$

Received 8 August; accepted 11 December 1996.

1. Bundy, F. P., Hall, H. T., Strong, H. M. & Wentorf, R. H. *Nature* **176**, 51–53 (1956).
2. Spitsyn, B. V., Bouilov, L. L. & Deryagin, B. V. *J. Cryst. Growth* **52**, 219–226 (1981).
3. Matsumoto, S., Sato, Y., Kamo, M. & Setaka, N. *Jpn. J. Appl. Phys.* **21**, L183–L185 (1982).
4. Shatsky, V. S. & Sobolev, N. V. *Dokl. Akad. Nauk SSSR* **331**(2), 217–219 (1993) (Russ.).
5. DeVries, R. C. in *Advanced Ceramics III* (ed. Somiya, S.) 181–205 (Elsevier Applied Science, London, 1990).
6. DeVries, R. C., Roy, R., Somiya, S. & Yamada, S. *Trans. Mater. Res. Soc. Jpn.* **14B**, 1421–1445 (1994).
7. Roy, R. *et al. Innov. Mater. Res.* **1**, 65–87 (1996).
8. Patel, A. R. & Cheria, K. A. *J. Cryst. Growth* **46**, 706–708 (1979).
9. Roy, R., DeVries, R. C., Ravichandran, D., Cheria, K. A. & Badzian, A. *Proc. 4th Int. Symp. on Diamond Mater.* (Electrochem. Soc., Pennington, NJ, USA, 1995).
10. Zhao, X.-Z., Cheria, K. A., Badzian, A. & Roy, R. Abstracts Am. Ceram. Soc. Meeting, Paper SXII-22-96 (Am. Ceram. Soc., Westerville, OH, USA, 1996).
11. Roy, R., Ravichandran, D., Badzian, A. & Breval, E. *Diamond Relat. Mater.* **5**, 973–976 (1996).
12. Roy, R., Ravichandran, D., Ravindranathan, P. & Badzian, A. *J. Mater. Res.* **11**, 1164–1168 (1996).
13. Szymanski, A. *et al. Diamond Relat. Mater.* **4**, 234–235 (1995).
14. Gogotzi, Y. G., Nickel, K. G. & Kofstad, P. *J. Mater. Chem.* **5**, 2313–2314 (1995).
15. Zhao, X.-Z., Cheria, K. A., Roy, R. & White, W. B. Abstracts Am. Ceram. Soc. Meeting, Paper SF-33-96 (Am. Ceram. Soc., Westerville, OH, USA, 1996); *J. Mater. Res.* (submitted).
16. Grot, R. C., Hatfield, C. W., Gildenblat, G. S., Badzian, A. R. & Badzian, T. in *New Diamond Science and Technology* (eds Messier, R., Glass, J. T., Butler, J. E. & Roy, R.) 949–955 (Mater. Res. Soc., Pittsburgh, USA, 1991).
17. Badzian, A. & Badzian, T. *Int. J. Refract. Met. Hard Mater.* (in the press).

Acknowledgements. This work was supported by ONR-ARPA. X.-Z.Z. thanks the NSF Division of Materials research for an SGER grant.

Correspondence and requests for materials should be addressed to R.R. (e-mail: rroy@psu.edu).

A decadal climate variation in the tropical Atlantic Ocean from thermodynamic air–sea interactions

Ping Chang, Link Ji & Hong Li

Department of Oceanography, Texas A&M University, College Station, Texas 77843–3160, USA

Rainfall variability in northeast South America¹ and the Sahel region of Africa^{2–4} is profoundly influenced by the sea surface temperature (SST) of the tropical Atlantic Ocean. Of particular importance are relative changes in SST between the hemispheres on decadal timescales, a phenomenon often called the Atlantic SST dipole^{1,5}. Here we propose that the decadal variation in the tropical SST dipole may be attributed to an unstable thermodynamic ocean–atmosphere interaction between wind-induced heat fluxes and SST. Using coupled ocean–atmosphere models, we show that the coupled dipole mode has a typical oscillation period of about a decade. The notion that the Atlantic dipole-like SST variability may be related to an oscillatory coupled mode might assist attempts to predict decadal climate variability in the tropical Atlantic region.

In some respects, the tropical Atlantic Ocean is similar to the tropical Pacific, having, for example, an equatorial cold tongue of sea surface temperature just south of the Equator, and the northeasterly and southeasterly trade winds converging in the Inter-tropical Convergence Zone (ITCZ) north of the Equator. The ITCZ and cold tongue complex in the Atlantic has a pronounced annual cycle, more prominent than that in the Pacific⁶. A phenomenon similar to but much weaker than the Pacific Ocean El Niño also occurs in the Atlantic⁷. A mode of variability seemingly unique to the tropical Atlantic is the SST dipole mode—the decadal variation of the interhemispheric SST gradient across the Equator. Studies have suggested that rainfall variabilities in the northeast Brazil and Sahel regions are closely related to the cross-equatorial SST gradient variation^{1–4}. In particular, the Brazilian rainfall tends to be more strongly correlated with the Atlantic SST dipole than with the Pacific El Niño, despite a stronger SST anomaly in the Pacific.

Dipole-like SST variabilities in the tropical Atlantic have been documented in several studies^{5,8,9}. We have applied a joint singular-value decomposition (SVD) analysis to a 40-year (1950–89) monthly mean surface data set that includes SST, surface wind-stress and heat-flux anomalies. This data set is derived from the reprocessed Comprehensive Oceans Atmosphere Data Set (COADS)¹⁰. The SVD analysis provides an efficient way to identify coupled patterns in the ocean–atmosphere system¹¹. Our analysis reveals that the first leading SVD mode, which explains about 45% of the total squared covariance, has a dipole-like SST pattern with the maximum amplitudes at about 15°S and 15°N, and opposite signs of response on either side of the Equator (Fig. 1a). A spectral analysis on the SST time coefficient of the first SVD (Fig. 1b) indicates a dominant spectral peak at about 13 years (Fig. 1c), consistent with previous studies^{8,9}.

Associated with the variability of the cross-equatorial SST gradient is a southeasterly wind anomaly in the Southern Hemisphere and a southwesterly wind anomaly in the Northern Hemisphere. These winds act to enhance the southeasterly mean trade winds in the Southern Hemisphere and reduce the northeasterly trades in the Northern Hemisphere. As a result, the surface heat flux from the ocean is decreased in the Northern Hemisphere the converse applies in the Southern Hemisphere. The anomalous heat flux enhances the

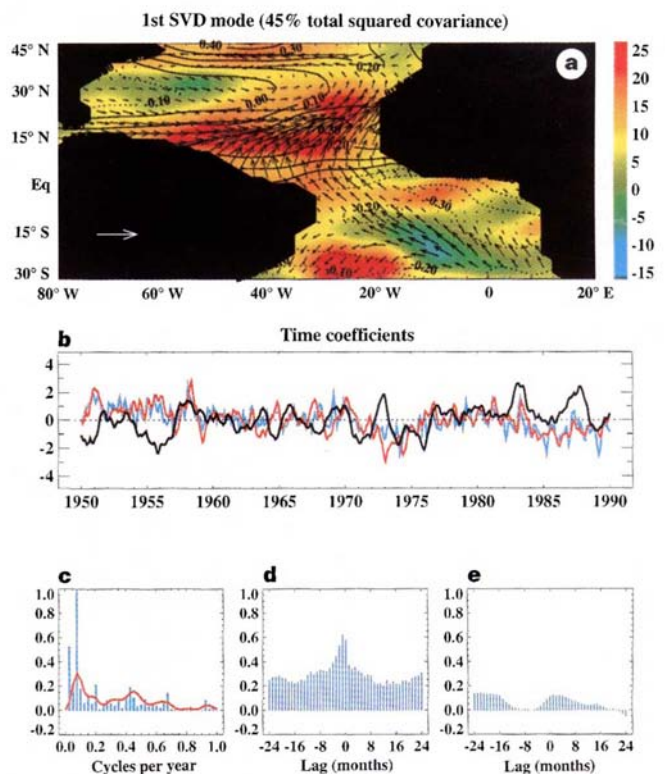


Figure 1 The Atlantic dipole-like variability revealed by a joint SVD analysis. **a**, Spatial structure of the first SVD mode. Contours represent SST anomaly in °C. Vectors depict wind-stress anomalies. The unit vector on the lower-left corner represents an easterly wind-stress anomaly of 1.0 dyn cm^{–2}. Colours indicate the strength of surface heat-flux anomalies in Wm². **b**, The two associated time coefficients (red line, 1st SVD of Atlantic SST; blue line, 1st SVD of Atlantic atmospheric variation). The time coefficients have been normalized by their own standard deviations. The black line in **b** is the normalized time coefficient of the first EOF of the global SST anomaly. **c**, Periodogram of the SST time coefficient, along with an estimated spectrum (red line) using the Hanning window with a bandwidth $M = 216$ months. **d**, Cross-correlation between the two time coefficients of the SVD, as a function of lag. The maximum correlation of 0.7 occurs at a lag of -1 , indicating that the atmosphere leads the ocean by one month. **e**, Cross-correlation between the time coefficients of the first SVD of the tropical Atlantic SST and the first EOF of the global SST. Low correlation values in **e** suggest that the coupled dipole mode is not strongly affected by changes in the global SST.

SST dipole. The high correlation between the surface heat flux and SST anomalies (SSTA) is illustrated in Fig. 1a and suggests a positive ocean–atmosphere feedback through thermodynamic processes. This positive feedback differs considerably from the El Niño/Southern Oscillation (ENSO) mechanism, where ocean dynamics (changes in equatorial upwelling and thermocline depth in response to anomalous winds) play a central role in regulating tropical SSTs. Carton *et al.*¹² and Zhou¹³ demonstrated that wind-induced latent heat flux is a major contributing factor to the variability of the SST dipole, whereas wind forcing is the main source of the interannual SST variability associated with the ‘Atlantic El Niño’.

Information on causal linkages of air–sea feedbacks may be contained in the cross-correlation between the time coefficients of the atmospheric forcing and SST¹⁴. A positive feedback implies that the ocean and atmosphere are mutually reinforcing each other. The cross-correlation in this case will be more or less symmetrical and of the same sign for both positive and negative lags, although it usually peaks when the atmosphere leads the ocean. In contrast, when the atmospheric forcing has negative or no feedbacks, the cross-correlation will have either an anti-symmetric appearance or a very asymmetric form, peaking when the atmosphere leads but dropping

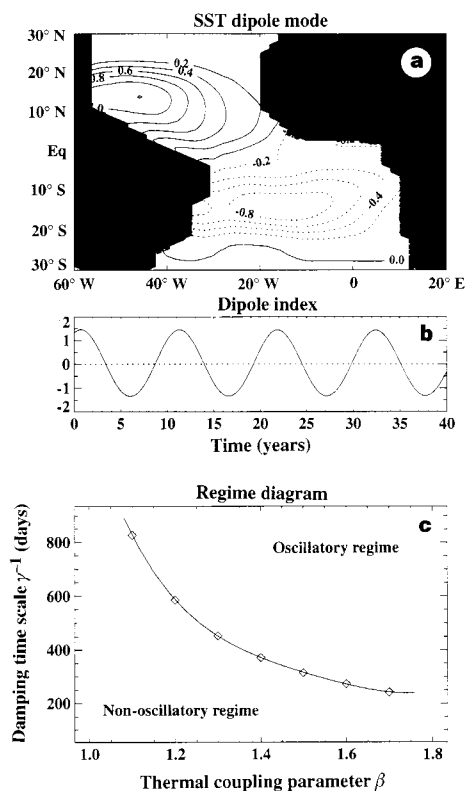


Figure 2 Dipole modes of the intermediate coupled model. **a**, Spatial pattern obtained by regressing the simulated SST upon a dipole index (shown in **b**) generated by differencing area-averaged SSTs in the Northern and Southern hemispheres. The area-averages were taken over longitudinal bands 18° wide centred at 15°S and 15°N across the basin. The results shown in **a** and **b** were based on simulation using coupling parameters $\alpha = 0$ and $\beta = 1.3$. The standard deviation of the dipole index is about 1.0°C. Contour intervals are 0.2°C per standard deviation of the dipole index. The regime diagram at the bottom was obtained by varying the thermodynamic coupling parameter β and damping parameter γ . The damping coefficient γ^{-1} determines the time over which the model SSTs are relaxed back to the annual mean SST by means of a newtonian damping term $-\gamma(T - T_{\text{annual}})$ in the model temperature equations. A boundary in parameter space separates the non-oscillatory regime from the oscillatory regime.

rapidly to zero when the atmosphere lags. Figure 1d shows the cross-correlation for the first SVD mode. The cross-correlation shows a peak when the atmosphere is leading the ocean by about a month and remains positive for all the lags, suggesting a positive feedback. Figure 1e illustrates the correlation between the dipole variation and global SST fluctuation represented by the time coefficient of the first leading empirical-orthogonal-function (EOF) mode of the global SST anomaly (black line in Fig. 1b). There is a low correlation (less than 0.2) throughout ± 24 -month lags, suggesting that the Atlantic dipole mode involves mainly local dynamics and is not strongly influenced by global processes.

To explore further the role of air–sea interactions in the tropical Atlantic, we have developed two coupled ocean–atmosphere models. The simpler model is an intermediate coupled model (ICM), which consists of an empirical atmospheric feedback model and a reduced-gravity ocean model^{15,16}. The more sophisticated model—a hybrid coupled general circulation model (HGCM)—combines the same empirical atmosphere with a 3-D primitive-equation Ocean General Circulation Model (OGCM; R. C. Pacanowski, K. Dixon and A. Rosati, personal communication). The model configurations are similar to those developed for

the Pacific ENSO studies described in ref. 17. The atmospheric feedback is constructed using the joint SVD analysis described earlier, which establishes an empirical relation between the SST anomaly and the corresponding surface wind-stress and heat-flux anomalies from observations of low-frequency variability¹⁸. Two scaling factors α and β are introduced for, respectively, surface wind stress and heat flux. These coupling parameters control the strength of dynamic and thermodynamic feedbacks. In the coupled models, only the spatial patterns associated with the singular eigenvectors are taken into consideration—that is, the SSTA pattern determines the patterns of wind stress and heat flux. Temporal information in the observations is not used in the coupled models.

A large number of numerical experiments were conducted with the coupled ocean–atmosphere models. For the ICM, each experiment consists of a 5-year spin-up forced with the annual mean wind stress and surface heat flux followed by a 60-year coupled integration. The HGCM experiments use a 120-year spin-up and a 120-year coupled run, as a fully stratified ocean model needs a much longer spin-up time. In all the experiments, the coupled models are forced externally with the COADS annual mean wind stress and heat flux. Figure 2 summarizes the results from a set of ICM experiments where the atmosphere and ocean are coupled only through thermodynamic processes; feedbacks between winds and SST are disallowed by setting α to zero. The spatial pattern of the coupled mode bears a striking resemblance to the observed SST dipole, with an oscillation period varying from 8 to 14 years depending on the thermodynamic coupling strength β . The larger β , the shorter is the oscillation period. When only the dynamic coupling is allowed ($\beta = 0$), the model produces an ENSO-like variability with oscillation period ranging from 2.5 to 4 years (not shown). This result is consistent with previous findings⁷. Figure 2c illustrates the dependence of oscillatory behaviour of the coupled modes on coupling and damping parameters. The HGCM experiments gave a similar result. A self-sustained dipole oscillation was obtained, when $\alpha = 0$, $\beta = 1.05$ and damping coefficient $\gamma^{-1} = 150$ days, with a period of about 13 years, indicating that the existence of the coupled dipole mode is model-independent.

To gain further understanding of the physical processes that control the dipole-like oscillations, we have analysed the evolution of model SSTs by diagnosing the upper-ocean heat budget. For a dipole-like oscillation, the dominant physical processes include a positive feedback between the heat-flux and SST anomalies, and a negative feedback due to the advection of heat by steady ocean currents. The negative feedback can be briefly described as follows: if the Northern Hemisphere SST is warmer than the Southern Hemisphere, the steady ocean currents, particularly the intensified western boundary currents such as the North and South Brazil currents, will transport anomalous cold (warm) water to the Northern (Southern) Hemisphere. This horizontal heat advection acts against the atmospheric warming (cooling) in the Northern (Southern) Hemisphere by transporting the anomalously cold (warm) water to the dipole region. Atmospheric radiative and turbulent energy fluxes, together with ocean mixing processes, damp SST anomalies, acting as additional negative feedbacks. When these positive and negative feedbacks properly balance each other, the system gives rise to a self-sustained oscillation. For the dipole oscillation, we have found that thermocline variability does not contribute significantly to balancing the heat budget, indicating that ocean wave dynamics may be of secondary importance. Thus, this coupled mode can be regarded as a generalization of the SST mode studied by Neelin¹⁹.

Although simple coupled models suggest that a self-sustained oscillatory dipole mode is possible in a coupled system, in reality stochastic processes generated by internal dynamics of the atmosphere and oceans can have a large impact on the response of

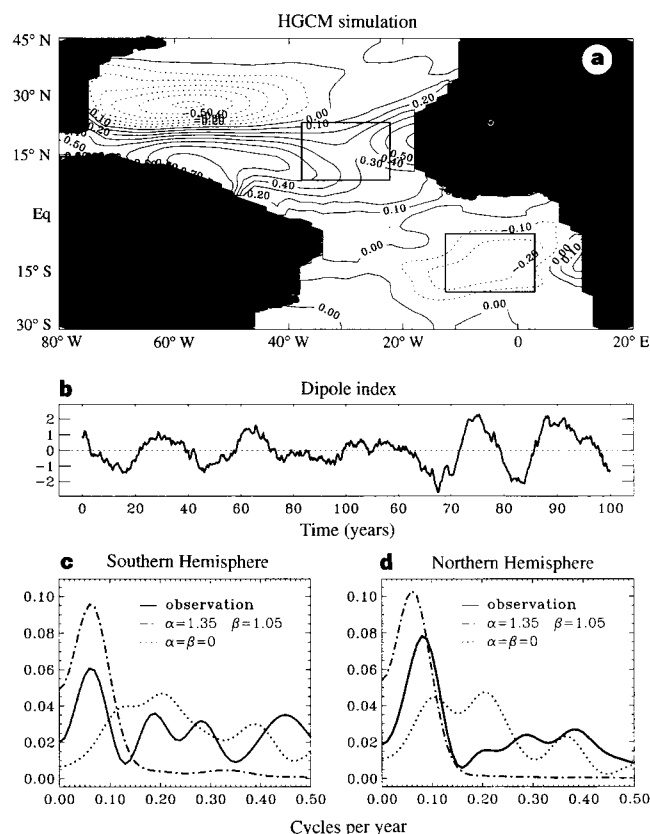


Figure 3 SST fields simulated by the hybrid coupled general circulation model forced with wind noise with both dynamic and thermodynamic coupling ($\alpha = 1.35$, $\beta = 1.05$). **a**, Dipole pattern generated using a regression analysis as in Fig. 2, with a dipole index (shown in **b**) derived by taking the difference of the model SSTs averaged over a $15^\circ \times 15^\circ$ area in each hemisphere indicated by two rectangles. The dipole pattern is insensitive to the choice of the index. The two areas were chosen because they had the best availability of 100-year SST observations. The 100-year (1890–1990) observed monthly mean SST time series were derived from the UK Meteorological Office Main Marine Data Bank²⁰ by area-averaging over the two regions. The observed SSTs were compared with similar SST time series derived from model simulations via a spectral analysis. **c**, **d**, SST spectra in the Southern and Northern hemispheres, respectively, where solid lines are for observations, dash-dotted and dotted lines are for simulations with $\alpha = 1.35$, $\beta = 1.05$ and $\alpha = \beta = 0$, respectively. The SST spectra were estimated using the Hanning window with a bandwidth $M = 240$ months and were normalized by their own variance. The standard deviations of the observed SST in the Northern and Southern hemispheres are about 0.5°C . Similar values for the simulated SSTs are about 0.3°C for $\alpha = 1.35$, $\beta = 1.05$ and are less than 0.1°C for $\alpha = \beta = 0$.

ocean–atmosphere coupled systems. Recent studies¹⁶ suggest that random atmospheric forcing can resonantly excite coupled modes in a system that contains no self-sustained oscillatory modes in the absence of stochastic processes. To explore the effects of stochastic processes on tropical Atlantic climate variability, we have done experiments with the addition of external atmospheric noise forcing. Figure 3 depicts the SST response from an HGCM experiment using $\alpha = 1.35$ and $\beta = 1.05$ with wind-noise forcing. The wind noise was constructed using the high-frequency component of the observed wind stress from COADS. It retains a coherent structure in space that mimics short period ‘weather’, but has a white spectrum in time. The coupled model produces dipole-like SST variability similar to observations, although the SST variabilities in the model are weaker than observed values (Fig. 3). A comparison between the observed and simulated SST spectra indicates that the coupled model reproduces the observed decadal spectral peak at a period around 13 years in both hemispheres (Fig.

3c and d). In contrast, the model SST response in the absence of any air–sea feedbacks ($\alpha = \beta = 0$) produces neither dipole-like SST patterns nor pronounced decadal spectral peaks. Experiments using white spatiotemporal wind and heat-flux noises lead to similar results, indicating that ocean–atmosphere interactions are essential to the decadal dipole-like SST variability in the tropical Atlantic Ocean.

In light of these results, we propose that the dipole-like variability in the tropical Atlantic may be attributed to ocean–atmosphere interactions involving primarily positive thermodynamic feedbacks between surface heat flux and SST. The coupled model experiments indicate that in a realistic parameter regime the dipole mode oscillates on a timescale of 12–13 years. In reality, stochastic processes in the atmosphere may play an important role in resonantly exciting the dipole mode and causing irregularities in the dipole variability. Therefore, the decadal dipole-like SST variability are perhaps best viewed as a resonant response of the coupled system to short-period ‘weather’ disturbances in the atmosphere. □

Received 25 June; accepted 17 December 1996.

1. Moura, A. D. & Shukla, J. *J. Atmos. Sci.* **38**, 2653–2675 (1981).
2. Lamb, P. J. *Tellus A* **35**, 198–212 (1983).
3. Hastenrath, S. *Mon. Weath. Rev.* **112**, 1097–1107 (1984).
4. Folland, C., Palmer, T. & Parker, D. *Nature* **320**, 602–607 (1986).
5. Servain, J. *J. Geophys. Res.* **96**, 15137–15146 (1991).
6. Philander, S. G. *El Niño, La Niña, and the Southern Oscillation* (Academic, New York, 1990).
7. Zebiak, S. J. *J. Clim.* **6**, 1567–1586 (1993).
8. Mehta, V. M. & Delworth, T. J. *J. Clim.* **5**, 172–190 (1995).
9. Mann, M. E. & Park, J. *J. Geophys. Res.* **99**, 25819–25833 (1994).
10. da Silva, A., Young, A. C. & Levitus, S. *Atlas of Surface Marine Data 1994*, Vol. 1, *Algorithms and Procedures* (NOAA Atlas NESDIS 6) (US Dept Commerce, Washington DC, 1994).
11. Bretherton, C. S., Smith, C. & Wallace, J. M. *J. Clim.* **5**, 541–560 (1992).
12. Carton, J. A., Cao, X., Giese, B. S. & da Silva, A. *J. Phys. Oceanogr.* (in the press).
13. Zhou, Z. thesis, Univ. Maryland (1996).
14. Frankignoul, C. & Hasselmann, K. *Tellus* **29**, 284–305 (1977).
15. Zebiak, S. E. & Cane, M. A. *Mon. Weath. Rev.* **115**, 2262–2278 (1987).
16. Chang, P. *J. Geophys. Res.* **99**, 7725–7741 (1994).
17. Chang, P., Ji, L. & Flügel, M. *Physica D* (in the press).
18. Syu, H. H., Neelin, J. D. & Gutzler, D. *J. Clim.* **8**, 2121–2143 (1995).
19. Neelin, D. *J. Atmos. Sci.* **48**, 584–606 (1991).
20. Parker, D. E., Jones, P. D., Folland, C. K. & Bevan, A. *J. Geophys. Res.* **99**, 14373–14399 (1994).

Acknowledgements. We thank G. Philander, M. Latif, D. Battisti, S. Xie, M. Flügel and T. Crowley for comments on earlier versions of the manuscript and J. Lynne for proofreading the manuscript. This work was supported by the NOAA Climate and Global Change Program and the NSF Young Investigator Program.

Correspondence should be addressed to P.C. (e-mail: ping@shark.tamu.edu).

Triggering basaltic volcanic eruptions by bubble–melt separation

Andrew W. Woods* & Silvana S. S. Cardoso†

* School of Mathematics, University of Bristol, Bristol BS8 1TW, UK

† Department of Chemical Engineering, University of Cambridge, Pembroke Street, Cambridge CB2 3RA, UK

Understanding the processes by which volcanic eruptions are triggered is crucial for volcanic hazard prediction and assessment. In intermediate and silicic magmas, where water is the dominant volatile phase^{1,2}, ‘second boiling’ provides an effective eruption trigger¹—as the magma cools and crystallizes, volatiles are increasingly concentrated in the residual melt where they eventually become saturated and are exsolved, thereby raising the magma pressure to the point of eruption^{1,2}. In contrast, in basaltic magma chambers in which carbon dioxide is the dominant volatile phase, the effect of second boiling is to decrease the chamber pressure. But basaltic magmas are also relatively inviscid, so volatile bubbles can separate from the turbulently convecting melt to produce a foam layer at the chamber roof^{3,4}; decompression of the rising bubbles can increase the chamber pressure and so trigger an eruption^{5,6}. Here we model the complex competition between the processes of magma cooling and bubble

ascent. We find that in basalt of very low viscosity (1–10 Pa s), bubble separation dominates and can increase the reservoir pressure by as much as 10 MPa (sufficient to trigger an eruption⁷) on timescales of 1–10 years, comparable to the duration of eruptive episodes at Kilauea volcano, Hawaii^{4,8}.

Basaltic magma chambers are typically in a state of vigorous convection⁹ with the volatiles well mixed throughout the melt. The timescales of bubble ascent and of the convective plumes that carry the bubbles through the chamber are relatively long compared to the time for volatile–melt equilibration, and so the volatiles may be assumed to be in equilibrium with the melt^{10,11}. The saturation concentration of CO₂ volatiles in basaltic melt increases with pressure¹, p , and hence depth below the surface, h , according to

$$n_s(h) = sp = s \left(\int_0^h \rho g dz + \Delta p \right) \quad (1)$$

where Δp is the chamber overpressure relative to lithostatic, $\rho(z)$ is the density of the country rock a distance z below the surface and the saturation constant s has value $4 \times 10^{-11} \text{ Pa}^{-1}$. As the chamber walls are elastic, we assume that the overpressure has the same value relative to the lithostatic pressure at all depths. Thus, if the total concentration of volatiles in a well mixed melt–crystal suspension is n_t , then the mass of exsolved volatiles at each depth, $n_e(z)$, is $n_e(z) \approx n_t - n_s(z)(1-x)$ where x is the mass fraction of CO₂-free crystals.

If at some stage the foam at the top of the chamber has depth d and the total chamber depth is H , then conservation of the total mass of volatiles, M (per unit area), requires that

$$n_t \int_{d-L}^{H+L} \rho_m dz + \rho_g d = M \quad (2)$$

where L is the distance of the chamber roof below the Earth's surface, ρ_g is the density of the volatiles at the top of the chamber and ρ_m is the density of the melt–volatile mixture. The gas density may be approximated¹ by $\rho_g = p/RT$, and because the foam is thin, in this expression, we evaluate the pressure at the top of the chamber. At each depth, the density of the bubble–melt mixture, ρ_m , is given in terms of the partial volumes of melt and gas¹⁰

$$\frac{1}{\rho_m} = \frac{n_e RT}{p} + \frac{1-n_e}{\sigma} \quad (3)$$

where σ is the melt density. Owing to the bubbles, the mixture is much more compressible than the surrounding country rock so that, to leading order, the chamber volume remains fixed. Thus, for a given foam thickness d and initial chamber pressure $p_0(h)$, the chamber overpressure associated with the ascent and decompression of the bubbles may be found by combining equations (1)–(3).

As magma is emplaced in the chamber, a fraction of the volatiles are exsolved, producing a melt–bubble suspension. In the absence of significant cooling, these bubbles ascend to the roof. Eventually the melt at the base of the chamber ceases to be saturated. However, the convective mixing leads to further exsolution of volatiles as magma deep in the chamber is brought to shallower levels, and an unsaturated layer of magma grows above the base of the chamber (Fig. 1). Ultimately, the saturation surface may rise to the top of the melt leaving a body of unsaturated magma below the foam (Fig. 1). This limit of slow cooling provides an estimate of the maximum pressure increase of the chamber through bubble separation (Fig. 2). In relatively shallow chambers, this overpressure may increase to values of the order of 10 MPa, which is sufficient to trigger eruption^{1,5,7}. We now determine conditions under which the bubbles can escape to the top of the chamber before significant cooling and crystallization of the melt has occurred.

By analogy with sedimenting particles, the bubble concentration in a well mixed melt evolves according to the law^{12,13}

$$\frac{dn_e}{dt} = F - \frac{v_s n_e}{H-d} \quad (4)$$

where $v_s n_e/(H-d)$ is the bubble flux supplied to the foam per unit mass, v_s is the Stokes rise speed of a bubble¹⁴, n_e is evaluated at the top of the chamber, and F is the rate of bubble production per unit mass as the melt cools and crystallizes. We have successfully tested equation (4) with a series of laboratory experiments in which bubbles were produced by electrolysis at the base of an experimental tank filled with aqueous saline solution¹⁵; these will be reported elsewhere.

The conservation of heat in a convecting melt is given by the approximate bulk heat balance for the chamber^{9,16}

$$\rho_m C_p (H-h) \frac{dT}{dt} = -F_T + \rho_m (H-h) L \frac{dx}{dt} \quad (5)$$

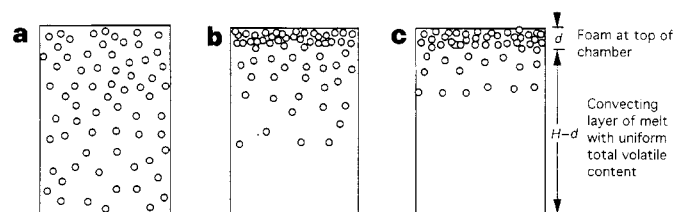


Figure 1 Diagrams of the ascent of bubbles through a well mixed layer of basaltic melt, illustrating how the melt becomes progressively unsaturated at the base of the chamber as the foam layer grows at the chamber roof. **a**, Initially there is a well mixed suspension of bubbles throughout the melt, although as the melt moves up or down, some volatiles are either exsolved or resorbed. As a result, in the deeper, high-pressure part of the chamber a larger fraction of the volatiles are dissolved in the magma. **b**, Bubbles separate into the upper layer of foam and so the volatile content of the melt decreases, until eventually the lower part of the chamber becomes unsaturated. **c**, Eventually many of the volatiles have ascended into the foam and only a small part of the melt layer contains exsolved volatile bubbles. Deeper in the chamber, the melt is unsaturated.

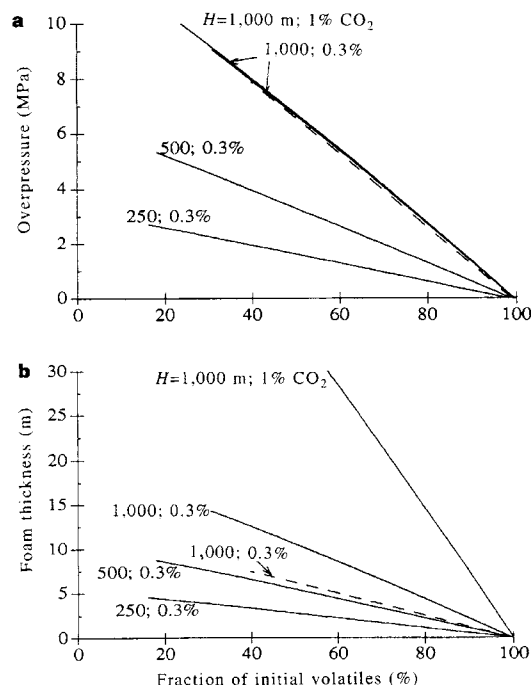


Figure 2 Increase of the chamber pressure (**a**) and the foam depth (**b**) as bubbles progressively migrate to the roof of the chamber. Curves are shown for chambers whose upper surface is 4 km below the surface (solid lines) and which have vertical extents of 1,000 m, 500 m and 250 m and for a chamber whose upper surface is 7 km below the surface (dashed line) and which has a vertical extent 1,000 m. For all but one curve, the magma has an initial volatile content of 0.3 wt%. For the final curve, corresponding to a chamber 4 km below the surface and of vertical extent 1,000 m, the magma has an initial volatile content of 1.0 wt%. These conditions are representative of many basaltic magmas^{4,5}.

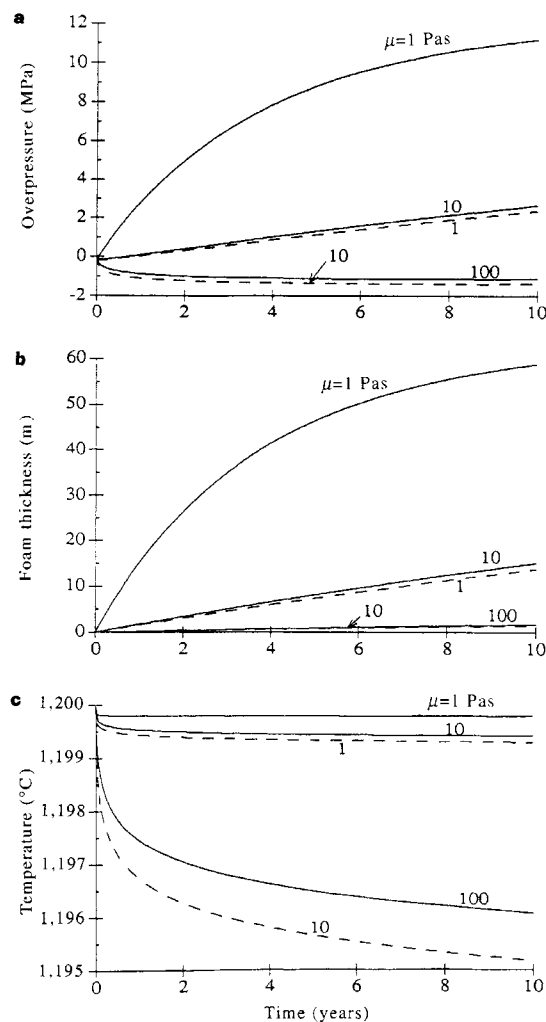


Figure 3 Variation with time of the chamber pressure (a), foam thickness (b) and melt temperature (c). Curves are shown for magma of initial viscosity (μ) of 1, 10 and 100 Pa s (as labelled on the curves) in a chamber of vertical extent 1 km whose upper surface is located 4 km below the surface. The solid lines correspond to bubbles of size 100 μm . The dashed lines correspond to bubbles of size 30 μm . In lower-viscosity melt, the foam develops rapidly, insulating the underlying magma, suppressing cooling and promoting further bubble separation. This causes a build-up of chamber pressure and the possible triggering of eruption. With more viscous magma, fewer bubbles can ascend to the top of the chamber in a given time, and so the magma cools and crystallizes more; the associated contraction of the melt may then lower the reservoir pressure. As with viscous magma, the smaller bubbles (30 μm) separate more slowly and so the effects of magma cooling become more important, again reducing the build-up of pressure.

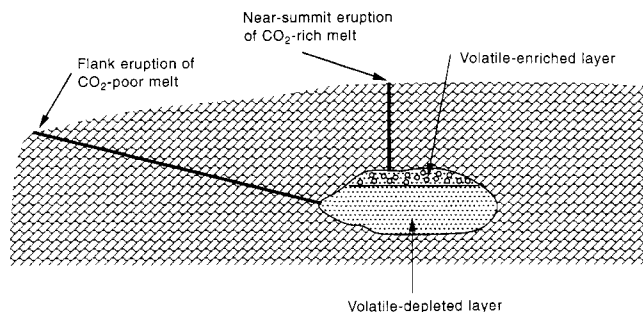


Figure 4 Sketch of a CO₂-poor eruption along a lateral dyke tapping the lower part of the chamber, and a CO₂-rich eruption along a dyke tapping the upper layer of the volatile stratified chamber.

where L is the heat of crystallization. We assume the heat transfer across the non-convecting foam layer, $K(T_i - T_c)/d$, equals the convective heat flux in the melt¹⁷ $F_T = 0.1\rho_m C_p \kappa (g\sigma\alpha \times (T - T_i)^4 / \kappa \mu)^{1/3}$, where K is the effective thermal conductivity of the foam¹⁸, T_c is the country rock temperature, T_i is the temperature at the top of the melt and α is the thermal expansion coefficient, κ the thermal diffusivity, μ the viscosity and C_p the specific heat of the melt⁹. Finally, we parametrize the crystal fraction, x , as a function of melt temperature⁹, $x(T) = (7,200/T) - 6$, and the melt viscosity, μ , as a function of crystal content⁹, $\mu(x) = \mu_0(1 - 1.67x)^{-2.5}$.

Combining equations (1)–(5) we can calculate the chamber pressure, foam thickness and melt temperature as functions of time, assuming that the bubbles have typical size^{4,10} 30–100 μm at the chamber roof. As anticipated, we find the melt evolution is a very nonlinear process (Fig. 3) which is highly sensitive to the ratio of (1) the initial timescale for cooling of the melt, $\tau_c \approx 10H[\kappa\mu/g\sigma\alpha(T - T_c)]^{1/3}/\kappa$ and (2) the timescale for accumulation of the bubbles at the chamber roof, $\tau_b \approx H/v_s \approx H\mu/g\sigma r_b^2$ where r_b is the bubble radius.

If the melt is relatively inviscid or the bubbles are relatively large, then $\tau_b < \tau_c$ and bubbles can form a small layer of foam before there has been significant cooling of the melt (for example, Fig. 3, curves with $\mu = 1, 10$ Pa s or $d = 100 \mu\text{m}$). This foam insulates the melt which then remains nearly isothermal. As bubbles continue to accumulate in the foam, the chamber pressure builds up. Eventually an eruption of very hot, low-viscosity basalt may be triggered. If the melt is more viscous or the bubbles are smaller, then $\tau_b > \tau_c$ (for example, Fig. 3, curves $\mu = 100$). Now the melt can cool significantly while very little foam accumulates. This cooling increases the viscosity of the melt and suppresses the accumulation of foam. As a result, the chamber pressure decreases through the cooling and crystallization of the melt, and no eruption ensues.

The timescale of 1–10 years required for bubble–melt separation in low-viscosity basalt (Fig. 3) is consistent with the duration of historical eruption episodes in Hawaii^{3,8,19,20}, where very hot, low-viscosity basalt is erupted. Once a conduit has become established, continued bubble–melt separation in the chamber can provide the driving pressure to maintain eruptive activity until much of the CO₂ has degassed from the magma. Although the dynamics of CO₂ bubbles in the chamber may trigger an eruption, the actual style of venting can also be controlled by the exsolution of H₂O which occurs at much lower pressures, as the magma approaches the surface^{20,21}. However, in a water-poor basalt, magma driven from the lower, CO₂-depleted part of the chamber along a sub-horizontal dyke may produce effusive activity some distance from the chamber (Fig. 4), while magma tapped from near the upper, CO₂-rich part of the chamber may lead to explosive fire-fountaining. □

Received 31 October; accepted 31 December 1996.

1. Tait, S., Jaupart, C. & Vergnolle, S. *Earth Planet. Sci. Lett.* **92**, 107–123 (1989).
2. Blake, S. J. *Geophys. Res.* **89**, 8237–8244 (1984).
3. Jaupart, C. & Vergnolle, S. *Nature* **331**, 58–60 (1988).
4. Jaupart, C. & Vergnolle, S. *J. Fluid Mech.* **203**, 347–380 (1989).
5. Pyle, D. M. & Pyle, D. L. *J. Volcanol. Geotherm. Res.* **67**, 227–232 (1995).
6. Steinberg, G. S., Steinberg, A. S. & Merzhanov, A. G. *Mod. Geol.* **13**, 257–266 (1989).
7. Touloukian, Y. S., Judd, W. R. & Roy, R. F. *Physical Properties of Rocks and Minerals* (McGraw-Hill, New York, 1980).
8. Swanson, D. A., Duffield, W. A., Jackson, D. B. & Peterson, D. W. *US Geol. Surv. Prof. Pap.* **1956**, 1–59 (1979).
9. Huppert, H. E. & Sparks, R. S. J. *J. Petrol.* **29**, 599–624 (1988).
10. Sparks, R. S. J. *J. Volcanol. Geotherm. Res.* **3**, 599–624 (1988).
11. Hurwitz, S. & Navon, O. *Earth Planet. Sci. Lett.* **122**, 267–280 (1994).
12. Martin, D. & Nokes, I. *Nature* **332**, 534–536 (1988).
13. Sparks, R. S. J., Huppert, H. E., Koyaguchi, T. & Hallworth, M. *Nature* **361**, 246–249 (1993).
14. Batchelor, G. K. *An Introduction to Fluid Mechanics* (Cambridge Univ. Press, 1969).
15. Kimura, R. *Fluid Dyn. Res.* **3**, 395–399 (1988).
16. Jarvis, R. A. & Woods, A. W. J. *Fluid Mech.* **273**, 83–107 (1994).
17. Denton, R. A. & Wood, I. R. *Int. J. Heat Mass Transf.* **22**, 1339–1346 (1979).
18. Gebhart, B. *Heat Conduction and Mass Diffusion* (McGraw-Hill, New York, 1979).
19. Parfitt, E. A. *J. Geophys. Res.* **96**, 10100–10112 (1991).
20. Parfitt, E. A. & Wilson, L. *Bull. Volcanol.* **57**, 440–450 (1995).
21. Mangan, M. T., Cashman, K. V. & Newman, S. *Geology* **21**, 157–160 (1993).

Acknowledgements. We thank L. Wilson for providing valuable contributions to this work, and D. Pyle for useful discussions. This work was supported by the Leverhulme Foundation.

Correspondence should be addressed to A.W.W. (e-mail: a.w.woods@bris.ac.uk).

Potential collapse of North Sea cod stocks

R. M. Cook*, A. Sinclair† & G. Stefánsson‡

* SOAEFD Marine Laboratory, PO Box 101 Victoria Road, Aberdeen AB11 9DB, UK

† Department of Fisheries and Oceans, Gulf Fisheries Centre, PO Box 5030, Moncton, New Brunswick, Canada

‡ Marine Research Institute, PO Box 1390, Skulagata 4, 121 Reykjavik, Iceland

In common with many fish stocks in the North Sea, cod are heavily exploited with as much as 60% of the fishable stock being removed annually¹. The International Council for the Exploration of the Sea (ICES), which advises fishery managers on the state of fish stocks in the northeast Atlantic, has recommended that exploitation rates be reduced considerably and immediately, in order to prevent further stock decline¹. We examine the most recent ICES assessment of the North Sea cod stock¹ to see if the present exploitation regimen is sustainable. Our analysis suggests that not only is the present exploitation rate unsustainable, but that even regimens close to the maximum sustainable yield may be potentially prone to risk. There is a need for swift and effective action to protect the stock and avoid the problems of the much publicized collapse of cod stocks off the coast of Atlantic Canada^{2,3}.

The sustainability of harvesting is largely determined by two factors, the relationship between the size of spawning stock (SSB) and the annual number of offspring (the recruits) produced, and the subsequent survival of the recruits on entering the fishery. This is illustrated in Fig. 1a which shows a theoretical stock-recruitment curve and a recruit survivorship line. Where the two lines intersect is an equilibrium point to which the population is attracted⁴. If the survivorship line lies above the stock-recruitment curve there is no non-zero equilibrium point and the population is attracted to the origin. The slope of the survivorship line is affected by the fishing mortality rate. The more heavily the stock is exploited, the steeper the slope. This line is also called the replacement line because it defines the survivorship needed to replace the spawning stock in the future.

Figure 1b shows the stock-recruit data for North Sea cod taken from the most recent assessment¹. Although the points are highly scattered, it has been shown that there is a detectable effect of stock size on recruitment⁵. We can take this analysis further by fitting a conventional stock-recruitment curve to the data. The fitted Shepherd curve⁶, in which we assume log-normal autocorrelated errors, is shown in Fig. 1b. Also shown is the replacement line which corresponds to the present level of exploitation. As can be seen, the replacement line is more or less coincident with the rising slope of the stock-recruitment curve. If this mortality is correctly estimated and fishing is maintained at this level, the stock would be expected to collapse towards the origin. It is worth noting that since 1984 the time series of SSB and recruitment values follows the downward slope as predicted by the theory. Furthermore, the values from the early years, when the fishing mortality rate was about half the present level¹, appear to cycle around a stable equilibrium point to the right of the maximum on the stock-recruitment curve. This too is consistent with the theory.

There can never be full certainty that the underlying stock-recruitment relationship is the driving force behind the data points because a similar effect might be seen if an external environmental effect (a regimen shift) were to drive down recruitment during some decades⁷. In the case when such effects are measurable, they can be quantified and used to estimate different stock-recruitment curves for each environment regimen⁸. In the case of the

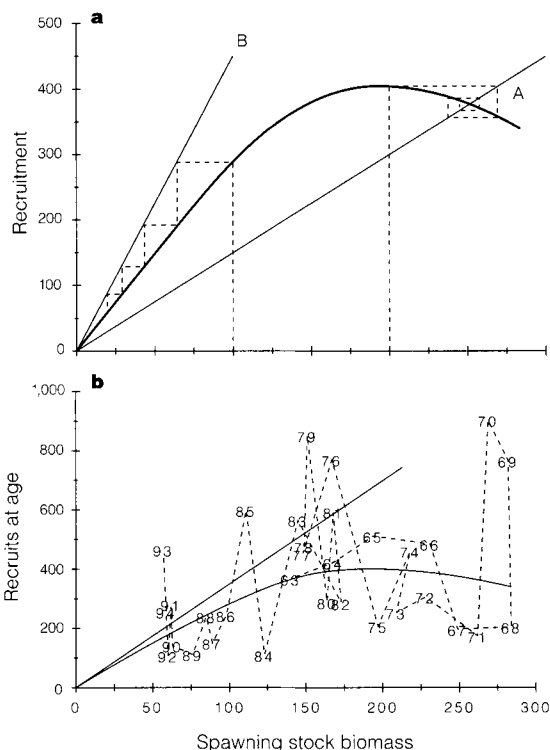


Figure 1 a. Theoretical stock-recruitment relationship for a fish stock. The curve predicts recruitment given stock size, and the straight lines are the replacement lines. These lines predict expected spawning stock from recruitment for two levels of exploitation. At a low exploitation rate (A), the dotted line indicates the population trajectory expected which cycles towards an equilibrium point. For high exploitation (B) the dotted line shows the expected population trajectory collapsing towards the origin. **b.** North Sea cod stock-recruitment data. The dotted line joins successive recruitment values with the year indicated. The straight line is the replacement line for the fishing mortality rate estimated for 1993. The fitted curve is a Shepherd stock-recruitment function.

North Sea cod, however, there is no known environmental factor or specific prey abundance that has been shown to affect recruitment in this manner.

Clearly, given the scatter of the data points, the parameters of the stock-recruitment curve are not well estimated and there must be doubt about the precise location of the curve in relation to the replacement line. If, however, we assume for the present that the fitted curve is an adequate description for the dependence of recruitment on SSB then it is possible to identify the conditions of sustainability. Figure 2a and b shows the relationship between yield, SSB and fishing mortality rate under equilibrium conditions using the stock-recruitment function fitted in Fig. 1b. Superimposed on these functions are the observed points joined in a time series. Both figures illustrate the problem of recent stock decline. Figure 2a shows the expected equilibrium SSB for a range of fishing mortality rates. The observed data follow the expected trajectory towards collapse with a fishing mortality rate of 0.9, close to the estimated present level¹. The observed yield tracks the expected yield in the same way (Fig. 2b). What is of particular concern is that the rate of decline predicted from the theoretical curve is very steep beyond the maximum sustainable yield (MSY) point and that the recent stock history is consistent with this prediction. It means that although MSY is a sustainable harvesting strategy, its proximity to conditions of stock collapse is sufficiently close to make management at MSY risky. This is because estimation errors will make it difficult to judge precisely where on the yield curve the present harvesting level is located.

The shape of the stock-recruitment function does much to determine the form of the equilibrium yield and SSB curves. In order to investigate the sensitivity of the predictions to the fitted curve, we also fitted Ricker⁹ and Beverton–Holt⁴ curves to the stock-recruitment data and calculated the fishing mortality at which the stock would be expected to collapse (Table 1). The values ranged from 0.91 to 1.13 compared with the estimated current fishing mortality of 0.91. This suggests that the conclusion that the stock is being fished at dangerous levels does not depend on the shape of the stock-recruitment function. The main effect of the choice of different recruitment curves is on the location of MSY. In the case of the Beverton–Holt curve, MSY is located at much lower fishing mortality rates and hence would be a safer exploitation strategy under this assumption. However, the estimated spawning stock at MSY is implausibly large, being a factor of six above the largest ever observed value and corresponds to a fishing mortality rate below any observed historical value.

A particular concern with cod is the pattern of exploitation with age. Cod can live many years and only reach maturity in significant numbers by the age of four. However, fish are caught as early as age one and by age two young fish are fully exploited by the fishery¹⁰. This means fish suffer substantial fishing mortality before they have a chance to reproduce. At present exploitation rates, only four per cent of fish aged one will survive to the age of four. This is the main reason for the steepness of the slope of the replacement line in Fig. 1b, which characterizes the low survival of recruits into the spawning stock.

Without a substantial reduction in the rate of fishing, the North Sea cod stock may well collapse. Overfishing has caused severe

Table 1 The results of fitting three different stock-recruitment functions to North Sea cod data

Function	Shepherd	Ricker	Beverton–Holt
Formula	$R = aS[1 + (S/b)^c]^{-1}$	$R = aS \exp(-bS)$	$R = aS/(1 + S/b)$
Parameter estimates (s.d.)	$a = 3.026 (0.523)$ $b = 248.72 (35.58)$ $c = 3.24 (1.958)$	$a = 4.151 (0.814)$ $b = 0.0039 (0.0011)$	$a = 4.900 (1.781)$ $b = 127.37 (86.30)$
Coefficient of determination	0.255	0.234	0.215
F_{crash}	0.91	1.05	1.13
F_{msy}	0.77	0.65	0.24
S_{msy}	203	298	1,736

R is recruitment and S is spawning-stock biomass. The quantities a , b and c are equation constants. (a is the slope of the curve at the origin, and the constants b and c determine the curvature in the stock-recruitment relationship). The standard deviation of the estimated parameters is given in parentheses. F_{crash} is the fishing mortality where the equilibrium spawning stock is zero. F_{msy} is the fishing mortality at maximum sustainable yield and S_{msy} is the equilibrium spawning stock at F_{msy} given in thousands of tonnes.

reductions in cod stocks in the northeast Arctic¹¹, Iceland¹² and Canada^{2,3}. Governments have taken severe management actions to help rebuild these stocks, by reducing the total allowable catch or closing fisheries completely. For many years, ICES has advised that the fishing mortality rate for North Sea cod should be reduced but data suggest management efforts have not been effective. There is now an urgent need to ensure that the exploitation rate is reduced either through the use of effective catch controls or by a direct reduction in fishing activity. □

Received 10 September; accepted 11 December 1996.

1. Report of the ICES Advisory Committee on Fishery Management. ICES Cooperative Research Report No. 214 (ICES, Copenhagen, 1995).
2. Hutchings, J. A. & Myers, R. A. *Can. J. Fish. Aquat. Sci.* **51**, 2126–2146 (1994).
3. Myers, R. A., Barrowman, N. J., Hoenig, J. M. & Qu, Z. *ICES J. Mar. Sci.* **53**, 629–640 (1996).
4. Beverton, R. J. H. & Holt, S. J. *Fishery Investigations Series* 2, 19 (HMSO, London, 1957).
5. Myers, R. A., Bridson, J. & Barrowman, N. J. *Canadian Technical Report of Fisheries and Aquatic Sciences* Report No. 2024 (1995).
6. Shepherd, J. G. J. *Cons. Intern. Explor. Mer* **40**, 67–75 (1982).
7. Cushing, D. H. J. *Cons. Intern. Explor. Mer* **41**, 159–166 (1984).
8. Jakobsen, J., Guðmundsdóttir, A. & Stefánsson, G. *Fish. Oceanogr.* **2**, 260–277 (1993).
9. Ricker, W. E. J. *Fish. Res. Board Can.* **11**, 559–623 (1954).
10. Cook, R. M., Kunzlik, P. A. & Fryer, R. J. *ICES J. Mar. Sci.* **48**, 1–13 (1991).
11. Jakobsen, T. *University of Alaska Sea Grant College Program Report* no. 93-02 (University of Alaska, Fairbanks, 1993).
12. Schopka, S. A. *ICES Mar. Sci. Symp.* **198**, 175–193 (1994).

Correspondence should be addressed to R.M.C. (e-mail: cookrm@marlab.ac.uk).

Extreme adaptive modification in sex ratio of the Seychelles warbler's eggs

Jan Komdeur^{*†‡}, Serge Daan^{*}, Joost Tinbergen^{*} & Christa Mateman[§]

^{*} Zoological Laboratory, University of Groningen, PO Box 14, 9750 AA Haren, The Netherlands

[†] National Environmental Research Institute, Department of Wildlife Ecology, Kalø, Grenåvej 12, DK-8410 Rønde, Denmark

[§] Netherlands Institute of Ecology, Centre for Terrestrial Ecology, PO Box 40, 6666 ZG Heteren, The Netherlands

Young Seychelles warblers *Acrocephalus sechellensis* often remain in their natal territories as helpers. Helpers on low-quality territories (as measured by food availability) reduce their parents' reproductive success, whereas 1–2 helpers on high-quality territories increase their parents' reproductive success, thereby

[‡] Present address: Department of Zoology, University of Melbourne, Parkville, Victoria 3052, Australia.

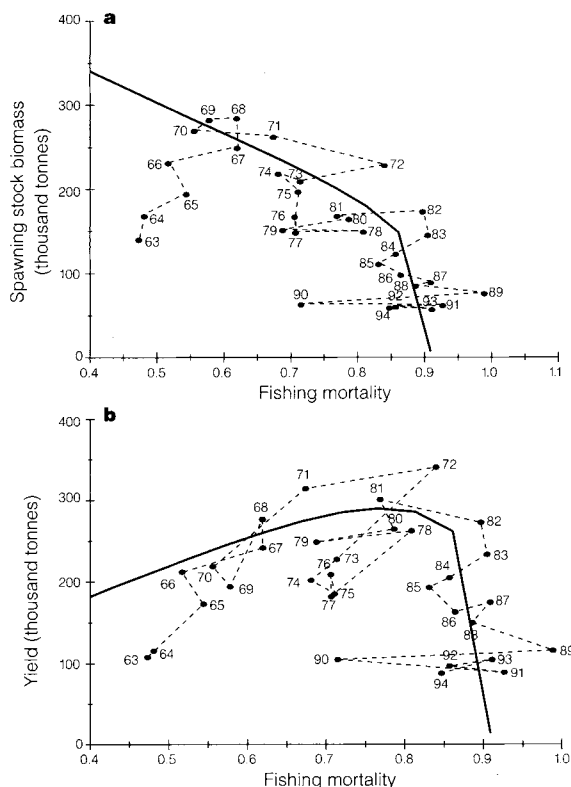


Figure 2 **a**, The equilibrium spawning stock biomass estimated from the stock-recruitment function fitted in Fig. 1b and the present exploitation pattern as a function of total fishing mortality. The plotted points are the observed values from the most recent assessment joined as a time series. **b**, The equilibrium yield estimated in the same way as **a**. The plotted points are the observed catches plotted as a time series. Both figures show the observed SSB and catch declining in line with the expected equilibrium function.

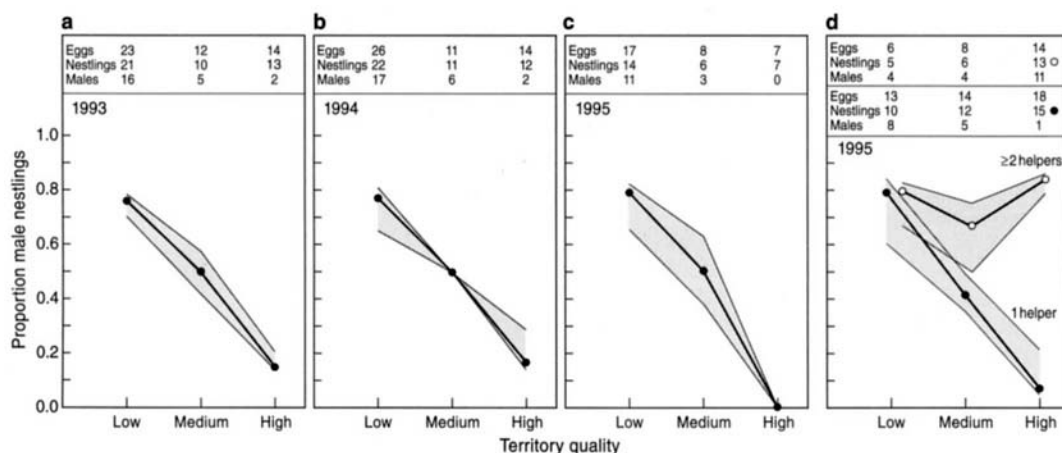


Figure 1 Sex ratio in Seychelles warbler nestlings. **a–c**, Sex ratio of nestlings produced by Seychelles warbler pairs in relation to quality class of breeding territory (*tg* classes: 1, low-quality territory; 2, medium-quality territory; and 3, high-quality territory; 1993–1995). No additional young were present on the territory. Young were hatched from one-egg clutches only in different years (**a**, 1993; *n* = 46, G-test of independence: $D = 12.23$, d.f. = 1, $P = 0.0005$, proportion male = $1/(1 + e^{-z})$, $z = -2.68 + 1.36(tg \text{ class})$; **b**, 1994: *n* = 45, $D = 12.03$, d.f. = 1, $P = 0.0005$, $z = -2.68 + 1.37(tg \text{ class})$; **c**, 1995: *n* = 27, $D = 12.99$, d.f. = 1, $P = 0.0003$, $z = -3.60 + 2.10(tg \text{ class})$), if in the analysis only young were included that had been hatched from different breeding pairs and different

mothers, the pattern of sex ratio of nestlings in relation to territory quality class remained the same (1993: *n* = 44, $D = 12.71$, d.f. = 1, $P = 0.0005$; 1994: *n* = 34, $D = 11.71$, d.f. = 1, $P = 0.0006$; 1995: *n* = 27, $D = 12.99$, d.f. = 1, $P = 0.0003$). **d**, Sex ratio of nestlings produced by Seychelles warbler pairs in relation to quality of breeding territory and to the number of helpers present (1995). ($D(tg \text{ class}) = 13.26$, d.f. = 1, $P = 0.0003$; $D(helper) = 10.86$, d.f. = 1, $P = 0.001$; $D(tg \text{ class} \times helper) = 11.77$, d.f. = 1, $P = 0.0006$; $z = -4.77 + 3.09(tg \text{ class}) + 1.93(helper) - 1.59(tg \text{ class} \times helper)$). Shaded area represents the maximal and minimal values for the sex ratio assuming that all eggs hatched were male, or females, respectively.

enhancing their inclusive fitness, in addition to gaining experience^{1,2}, and opportunities for co-breeding³. Helpers are mostly females, and we have previously suggested that parents may adjust the sex of their single egg to territory quality⁴. We therefore took blood samples from nestlings, and determined sex using random amplified polymorphic DNA (RAPD) markers. We show that biased hatching sex ratios are caused by biased production and not by differential embryo mortality. Unhelped breeding pairs on low-quality territories produce 77% sons, whereas unhelped pairs on high-quality territories produce 13% sons. Breeding pairs that were transferred from low- to high-quality territories switched from the production of male to female eggs. Breeding pairs occupying high-quality territories switched from producing female eggs when no or one helper was present, to producing male eggs when two helpers were present in the territory.

Until 1988, the world population of the endemic Seychelles warbler was entirely confined to Cousin Island (29 ha), where it has reached carrying capacity of about 320 birds^{5,6}. The warbler is insectivorous, gleaning insect food from leaves, and usually has a one-egg clutch (91.0%, *n* = 223) once per year and has a high annual adult survival (81.1%, 334 bird-years). The breeding pair remains in the same territory, sometimes for as long as nine years. Although warblers can breed successfully in their first year, some individuals remain on their natal territories as helpers, providing nourishment to their parents' offspring^{1–4}. Helpers are mainly females (88%, *n* = 271), which remain significantly longer in their natal territories than males (3.3 and 1.2 years, respectively)⁴. The frequency of helping is affected by habitat saturation and variation in territory quality^{5,6}, measured in terms of insect prey density. Both territory quality and the presence of helpers are important factors affecting the fitness of parents. Breeding pairs that were transferred from low- to high-quality territories improved their reproductive success significantly⁷. Removal experiments showed that the presence of 1–2 helpers in high-quality territories enhanced the reproductive success of their parents, due to helping behaviour, whereas the presence of helpers on low-quality territories and the presence of 3 or more helpers on high-quality

territories decreased future reproductive output of their parents owing to food competition¹. Here we assess whether warblers are capable of adjusting the sex of their eggs to the quality of the territory they inhabit and, in addition, to the number of helpers already present on the natal territory.

In the three years 1993–95, the fraction of male nestlings produced by unhelped breeding pairs changed significantly with territory quality (Fig. 1a–c). On average, most of the nestlings in low-quality territories were males (77.2%, *n* = 57). Neither sex predominated significantly in medium-quality territories (55.2% males, *n* = 27), and nestlings in high-quality territories were mainly females (87.5%, *n* = 32). The observed hatching sex ratio is not caused by differential embryo mortality. Even if we assume that the few unhatched eggs were of the minority sex in each case, a significant decrease in male nestlings with territory quality still remains (Fig. 1a–c).

The translocation of Seychelles warblers to the previously unoccupied islands of Aride (68 ha) and Cousine (26 ha) in September 1988 and in June 1990, respectively^{8,9} allowed us to test experimentally whether individuals indeed modify sex ratio of their clutch in relation to territory quality. Breeding pairs that were transferred from low-quality territories on Cousin to high-quality territories on the new islands, switched from a sex ratio biased towards male eggs before the transfer to mainly female eggs after the transfer (Table 1). Sex ratio of eggs laid by pairs breeding on high-quality territories before and after translocation were biased towards female eggs during both periods (Table 1).

Nestling sex ratio for pairs with one helper was not significantly different from that of pairs with no helpers, and again was negatively associated with territory quality (Fig. 1d). In contrast, pairs with two or more helpers produced mainly sons, even in territories with a high habitat quality (fraction of males, 0.85 (*n* = 13); Fig. 1d). Even if we assume that the unhatched eggs were of the same sex, the biased hatching sex ratio still remains (Fig. 1d). The sequence of egg sex ratio in successive broods in three consecutive years of the same pair in relation to the number of helpers present differed with habitat quality. The production of excess male eggs by pairs on low-quality territories was independent of the number of helpers present

Table 1 Comparison of sex ratios of eggs and territory quality (*tq*) class of Seychelles warbler pairs on Cousin Island before translocation, and of the same pairs on the islands of Aride and Cousine after translocation

<i>tq</i> class pair	Before translocation Cousin			After translocation Aride, Cousine		
	Low		χ^2 (d.f. = 3)	High		χ^2 (d.f. = 3)
	Daughters	Sons		Daughters	Sons	
1	0	6		12	0	
2	1	2		3	1	
3	1	6		9	2	
4	0	4		5	2	
Total	2	18	13.90**	29	5	18.76***

<i>tq</i> class pair	High		χ^2 (d.f. = 2)	High		χ^2 (d.f. = 2)
	Daughters	Sons		Daughters	Sons	
	Daughters	Sons		Daughters	Sons	
5	4	1		2	1	
6	6	2		8	1	
7	5	1		6	2	
Total	15	4	6.47*	16	4	8.11*

The difference between the changes in sex ratios of eggs of the two groups, originating from 'low- and high-*tq* classes was significant (G-test of independence: $D = 20.85$, d.f. = 1, $P < 0.0005$). * $P < 0.05$, ** $P < 0.025$, *** $P < 0.001$.

Table 2 The sex ratio (proportion male eggs) in successive clutches of the same pairs in relation to territory quality and number of helpers present

Year	1993	1994	1995	No. of	χ^2
Sequence of nestling production	First	Second	Third	pairs	(d.f. = 2)
With helpers in '94-'95					
Numbers of helpers	0	1	2		
Hatching sex ratio 1qt	1.00	1.00	0.80	5	2.05
Hatching sex ratio hqt	0.00	0.14	0.71	7	10.42*
Without helpers in '93-'95					
Number of helpers	0	0	0		
Hatching sex ratio 1qt	0.83	0.83	1.00	6	1.14
Hatching sex ratio hqt	0.20	0.00	0.00	5	2.05

All the eggs produced by these pairs resulted in a nestling, which was blood-sampled and sexed using RAPD markers. Abbreviations: 1qt: low-quality territory; hqt: high-quality territory. * $P < 0.01$.

(Table 2). In contrast, pairs on high-quality territories with no or one helper present produced mainly female eggs, but shifted sex ratio of their eggs to males when two helping offspring were present (Table 2). The third egg produced by pairs on high-quality territories was male when the first two offspring remained as helpers, but female when the first two offspring dispersed from the territory.

The removal of helpers on high-quality territories in June 1990¹ allowed us to test experimentally whether individuals indeed modify sex ratio of their clutch in relation to the number of helpers present. All the eggs laid by six groups during the breeding season before and during the breeding season after the removal of helpers produced nestlings, which survived to be blood-sampled in 1993 for RAPD sexing. Before helper removal, the six eggs laid by the six experimental units comprising the breeding pair with two helpers were all males; after removal of one helper, one son and five daughters were produced (Fisher exact test: $P = 0.007$). The sex ratio of the clutch laid by the reduced groups was exactly the same as that in six control units comprising the breeding pair and one helper (one son, five daughters, both in 1990 and 1991). The difference between the changes in sex ratio of eggs of the two groups, originally with different number of helpers present was significant (G-test of independence: $D = 5.98$, d.f. = 1, $P < 0.025$). Seychelles warblers adjust primary sex ratio facultatively in response to the quality of the territory they inhabit and to the number of helpers present on the breeding territories.

In contrast to the small variations in offspring sex ratios that seem to characterize most bird species¹⁰⁻¹⁴, the Seychelles warbler on Cousin Island shows extreme skews in hatching sex ratios. The sex ratio variation in Seychelles warblers offers support for local resource enhancement (LRE) through helping by one sex ('produc-

tion-of-helpers' hypothesis)^{11,15-18}. When helpers enhance the fitness of their parents, warbler breeding pairs produce predominantly the helping female sex. The sex ratio variation in the warblers is further consistent with the local resource competition (LRC) hypothesis^{17,18}. By biasing offspring sex ratio towards sons, which disperse, in low-quality territories and daughters in high-quality territories, breeding birds avoid having competing offspring on low-quality territories (LRC), and gain helpers on high-quality territories (LRE). With three or more helping birds present in a high-quality territory, the reproductive success of the parents decreases, resulting from increased risk of egg break caused by simultaneous incubation by four or more females, and greater depletion of food resources¹³. When two adult helping young were already present in high-quality territories, individual breeding pairs shifted from producing daughters to producing sons. Seychelles warblers increase their fitness by adaptively modifying the primary sex ratio. These results support the general proposition that sex ratio variation is expected when the profitabilities of raising sons or daughters vary between individuals¹⁹. To our knowledge, the Seychelles warbler is the first species in which future benefits accruing to the parents have been shown to contribute to the evolution of adaptive sex ratio manipulation. □

Methods

Data collection. On Cousin, all 115-123 breeding groups (310-400 birds) were checked regularly (every two weeks in 1985-1991; every four weeks in 1992-1996) for breeding activity from December 1985 to August 1996. Data were based on individually colour-ringed birds. As warblers have never colonized other islands by themselves, we assumed that missing birds had died if they were not found on other territories. Nestlings between 4 and 12 days old

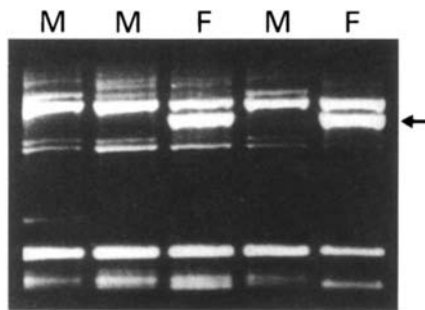


Figure 2 PCR products from genomic DNA of Seychelles warblers showing a female-specific fragment of 1.02 kb (indicated by an arrow; M, male; F, female).

and adults were blood-sampled in 1993–95 for sexing using RAPD markers. All nests were checked for the presence of a clutch, and clutch size during the 3–5 days after laying the first egg. In the analyses, only one-egg clutches were included, because these eggs were certainly from the breeding pair²⁰. From 1993 to 1995, all nestlings produced within a 5-week period, the main breeding season, by breeding pairs without helpers were sampled. Most of the nestlings born in the population hatched during this 5-week period (82.0%, $n = 273$)⁷. In 1995, all nests built within a 5-week period by all breeding groups, including those with helpers on Cousin were checked for a clutch, and all nestlings produced within this period were blood sampled.

A total of four breeding pairs from low-quality territories and three breeding pairs from high-quality territories on Cousin were transferred to Cousine and Aride^{8,9}. On the new islands they remained paired and occupied high-quality territories. All the eggs produced by these pairs pre- and post transfer resulted in a nestling, which survived to be blood-sampled in 1993 and sexed using RAPD markers.

Territory quality. Territory quality was expressed as mean insect prey available within a territory, because adult survival and reproductive success correlated with this^{3,6}. Territory quality was measured each month, and calculated as $tq = a \sum_{x=1}^{12} (c_x i_x) / 100 \Sigma$ where a is the mean yearly territory size (hectares), c_x is the mean yearly foliage cover for plant species x , and i_x is the mean monthly insect totals for plant species x per square decimetre of leaf area per year. Territories were divided into three classes: class 1 is low ($tq = 0–1,500$), class 2 is medium ($tq = 1,501–3,000$), and class 3 is high quality ($tq > 3,000$).

Sex determination. DNA was extracted²¹ from blood samples (10–50 μ l) collected from 4–12-day-old nestlings by brachial venipuncture. Amplification consisted of 38 cycles with primer 5'-GGGTAACGCC-3' according to ref. 22 and the products were separated on a 1.5% agarose gel. The PCR products from genomic DNA of all nestlings were scored blindly for the presence of the female-specific fragment by two uninformed persons (Fig. 2). In cases of disagreement, the analysis was repeated. All males ($n = 98$) and females ($n = 86$) seen copulating at a later stage were correctly sexed by this method.

Received 24 September; accepted 2 December 1996.

- Komdeur, J. *Behav. Ecol. Sociobiol.* **34**, 175–186 (1994).
- Komdeur, J. *Behav. Ecol.* **7**, 326–333 (1996).
- Komdeur, J. thesis, Univ. Cambridge (1991).
- Komdeur, J. *Proc. R. Soc. Lond. B* **263**, 661–666 (1996).
- Komdeur, J. *Nature* **358**, 493–495 (1992).
- Komdeur, J. *et al. Anim. Behav.* **49**, 695–708 (1995).
- Komdeur, J. *J. Biol. Rhythms* **11**, 333–346 (1996).
- Komdeur, J., Bullock, I. D. & Rands, M. R. *W. Bird Conserv. Int.* **1**, 177–185 (1991).
- Komdeur, J. *Biol. Conserv.* **76**, 143–152 (1994).
- Williams, G. C. *Proc. R. Soc. Lond. B* **205**, 567–580 (1979).
- Gowaty, P. A. & Lennartz, M. R. *Am. Nat.* **126**, 347–353 (1985).
- Clutton-Brock, T. H. *Ibis* **128**, 317–329 (1986).
- Gowaty, P. A. *Current Ornithology* Vol. 8 (ed. Power, D. M.) 141–172 (Plenum, New York, 1991).
- Daan, S., Dijkstra, C. & Weissing, F. J. *Behav. Ecol.* **7**, 426–430 (1996).
- Emlen, S. T., Emlen, J. M. & Levin, S. A. *Am. Nat.* **127**, 1–8 (1986).
- Lessells, C. M. & Avery, M. I. *Am. Nat.* **129**, 610–620 (1987).
- Clutton-Brock, T. H. & Iason, G. R. *Q. Rev. Biol.* **61**, 339–374 (1986).
- Johnson, C. N. *Nature* **332**, 726–728 (1988).
- Trivers, R. L. & Willard, D. E. *Science* **179**, 90–92 (1973).
- Kappe, A. L., Komdeur, J., Weising, F. L. & van de Zande, L. *Evolution* (submitted).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning. A Laboratory Manual*, 2nd edn (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989).
- Griffiths, R., Tiwari, B. *Proc. Natl Acad. Sci. USA* **90**, 8324–8326 (1993).

Acknowledgements. We thank L. Gabrielsen and A. Datema for their help in the field; C. Dijkstra for double-scoring of sexes of nestlings determined using RAPD markers; N. Davies, K. Lessells, T. Clutton-

Brock and P. Gowaty for comments on this manuscript; and the staff on the islands Cousin, Aride and Cousine for helping with the logistics of the project. This study was supported by grants from BirdLife International, from the Danish Natural Science Research Council, and the Netherlands Foundation for the Advancement of Tropical Research.

Correspondence and requests for materials should be addressed to J.K. (e-mail: J.Komdeur@zoology.u-nimelb.edu.au).

Female infertility in mice lacking connexin 37

Alexander M. Simon*, Daniel A. Goodenough†, En Li‡ & David L. Paul*

* Department of Neurobiology and † Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA

‡ Cardiovascular Research Center, Mass General Hospital East, Charlestown, Massachusetts 02129, USA

The signals regulating ovarian follicle development and the mechanisms by which they are communicated are largely undefined¹. At birth, the ovary contains primordial follicles consisting of meiotically arrested oocytes surrounded by a single layer of supporting (granulosa) cells. Periodically, subsets of primordial follicles undergo further development during which the oocyte increases in size and the granulosa cells proliferate, stratify and develop a fluid-filled antrum. After ovulation, oocytes resume meiosis and granulosa cells retained in the follicle differentiate into steroidogenic cells, forming the corpus luteum^{1,2}. It has been proposed that intercellular signalling through gap junction channels may influence aspects of follicular development^{3,4}. Gap junctions are aggregations of intercellular channels composed of connexins, a family of at least 13 related proteins that directly connect adjacent cells allowing the diffusional movement of ions, metabolites, and other potential signalling molecules⁵. Here we show that connexin 37 is present in gap junctions between oocyte and granulosa cells and that connexin-37-deficient mice lack mature (Graafian) follicles, fail to ovulate and develop numerous inappropriate corpora lutea. In addition, oocyte development arrests before meiotic competence is achieved. Thus, cell–cell signalling through intercellular channels critically regulates the highly coordinated set of cellular interactions required for successful oogenesis and ovulation.

Immunocytochemistry was performed on frozen sections of mouse ovaries to investigate the composition of oocyte–granulosa cell gap junctions. Antibodies specific for connexins (Cx) 37, 40 and 43 were used because their messenger RNAs had been previously detected in ovary^{6–8}. Anti-Cx37 antibodies⁹ labelled oocyte surfaces with a punctate staining pattern both in developing (Fig. 1a, b; 100 out of 100 cases) and in Graafian follicles (Fig. 1c, d; 30 out of 30 cases) but not primordial follicles (data not shown). This is consistent with the known location of oocyte–granulosa cell gap junctions³. Anti-Cx40 antibodies did not label ovarian follicles (data not shown). Antibodies to Cx43 labelled granulosa–granulosa gap junctions as expected⁶ but in contrast to anti-Cx37 antibodies, did not label the surface of the oocyte (Fig. 1g, h). Although low levels of Cx43 immunoreactivity in oocytes have been reported^{6,10}, our results indicate that Cx43 does not contribute significantly to oocyte–granulosa junctions and that Cx37 is the predominant oocyte connexin.

Uniform female infertility was observed when a targeted mutation in the Cx37 gene (Fig. 2a) was generated to test the role of intercellular communication in follicle development. Heterozygous (Cx37^{+/-}) mice, which appeared grossly normal, were interbred and the resulting litters genotyped by Southern blotting or polymerase chain reaction (PCR) (Fig. 2b). The ratio of genotypes was mendelian (1:2:1), indicating that there was no reduction in

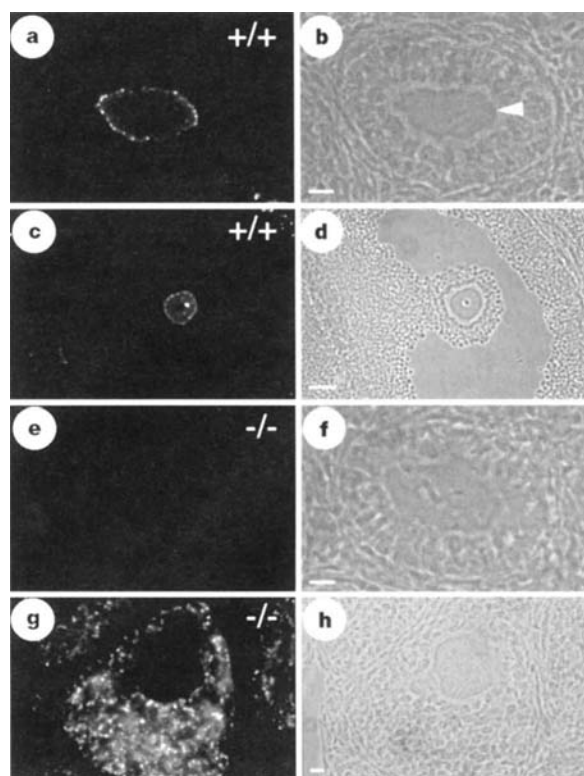


Figure 1 Connexin expression in ovarian follicles of wild-type and $Cx37^{-/-}$ mice. **a, b**, Punctate Cx37 immunostaining was observed at the oocyte surface (arrowhead), at the boundary between oocyte and zona pellucida, in preantral follicles. The location of Cx37 immunostaining is consistent with the known location of gap junctions formed between granulosa cell processes and the oocyte³. Faint and diffuse staining was observed within the ooplasm as well, presumably in intracellular membranes. **c, d**, Cx37 immunostaining was also found at the oocyte surface in mature Graafian follicles. **e, f**, As expected, oocytes from $Cx37^{-/-}$ mice showed no Cx37 immunostaining. **g, h**, Granulosa cells surrounding $Cx37^{-/-}$ oocytes displayed Cx43 immunostaining indistinguishable from wild-type controls⁸. Scale bars, 10 μ m (**a, b, e, f, g** and **h**); 50 μ m (**c, d**).

viability of homozygous ($Cx37^{-/-}$) mutant mice. Nine $Cx37^{-/-}$ females were mated at 5 weeks of age and produced no litters after being maintained in the presence of fertile males for >1 year. Although $Cx37^{-/-}$ mice showed no external abnormalities and appeared overtly healthy, $Cx37^{-/-}$ females were found to be completely incapable of ovulation, even in response to gonadotropin stimulation (data not shown). In contrast, $Cx37^{+/+}$ females showed normal fertility (20 females produced normal litters) as did both $Cx37^{+/+}$ and $Cx37^{-/-}$ males.

Oocytes in $Cx37^{-/-}$ animals were found to lack both recognizable gap junctions and detectable intercellular communication with granulosa cells. In wild-type follicles (Fig. 3a), granulosa cells adjacent to the oocyte extend processes through the zona pellucida forming both gap (Fig. 3c) and adherens junctions with the oocyte³. In $Cx37^{-/-}$ follicles (Fig. 3b), the zona pellucida and the granulosa cell processes within it were similar to wild-type in gross appearance and adherens junctions were still detected (Fig. 3d). However, neither oocyte–granulosa cell gap junctions (Fig. 3c) nor Cx37 (Fig. 1e, f) were observed in $Cx37^{-/-}$ follicles. Junctional communication was analysed by microinjection of neurobiotin. Injection of $Cx37^{+/+}$ oocytes resulted in transfer to granulosa cells in 10 out of 11 instances (Fig. 4a, b), whereas injection of $Cx37^{-/-}$ oocytes with neurobiotin resulted in no transfer to granulosa cells (8 injections) (Fig. 4c, d). Only oocyte–granulosa communication was inhibited in $Cx37^{-/-}$ follicles; granulosa cells directly injected

transferred neurobiotin readily to neighbouring granulosa cells (data not shown). Thus, infertility correlated with a loss of oocyte–granulosa communication.

Further analysis indicated that abnormalities in follicular growth, control of luteinization and oocyte maturation underlay $Cx37^{-/-}$ female infertility. Ovaries from $Cx37^{-/-}$ females displayed a profound defect in follicular growth (Figs 5 and 6). At 2 weeks after birth, ovaries of $Cx37^{-/-}$ females and wild-type females were similar in appearance (Fig. 5a). However, at sexual maturity (about 5 weeks), wild-type ovaries had surface distensions (arrowheads) caused by maturing follicles whereas $Cx37^{-/-}$ ovaries were smooth (Fig. 5b). This difference was even more pronounced in 4-month-old animals (Fig. 5c, d). The absence of surface distensions on $Cx37^{-/-}$ ovaries was explored by histological analysis (Fig. 6). At 2 weeks, both $Cx37^{-/-}$ and wild-type ovaries contained many developing preantral follicles (data not shown). By 5 weeks, however, wild-type ovaries routinely displayed numerous Graafian follicles (Fig. 6a, arrowheads) which were never observed in $Cx37^{-/-}$ ovaries (Fig. 6b). Occasional antral follicles were observed in the $Cx37^{-/-}$ ovaries that never grew larger than the example in Fig. 6b (arrowhead). The possibility that $Cx37^{-/-}$ animals could develop normal follicles with a slower time course seemed unlikely because Graafian follicles were still not detected by 4 months (Fig. 6d). Thus, follicular development consistently arrested before full maturation.

A second abnormality in $Cx37^{-/-}$ females was the inappropriate formation of corpora lutea. In normal animals, ovulation triggers mural granulosa to proliferate and differentiate into progesterone-secreting (luteal) cells, and promotes the invasion of blood vessels. Ovaries from wild-type mice displayed several corpora lutea about the size of Graafian follicles (Fig. 6c, arrowhead). $Cx37^{-/-}$ ovaries contained smaller structures resembling corpora lutea (Fig. 6d, arrowheads) that were 5–10 times more abundant. Electron microscopy (data not shown) verified that these structures displayed all of the characteristics of functional corpora lutea: high intracellular lipid content, tubular mitochondrial cristae, abundant smooth endoplasmic reticulum and numerous capillaries¹¹. Because luteinization in normal animals occurs following ovulation when, by definition, oocyte–granulosa cell communication is lost, the premature luteinization observed in $Cx37^{-/-}$ females suggests a model in which junctional communication is a major mechanism regulating corpus luteum formation. Consistent with this idea, surgical removal of oocytes from rabbit Graafian follicles causes morphological luteinization¹². Furthermore, our model predicts that an inhibitory signal is transferred from the oocyte through gap junctions to the granulosa cells and results in the prevention of corpus luteum formation.

In addition to inhibiting follicular growth and promoting luteinization, lack of junctional communication appears to influence oocyte development. In normal animals, oocytes are arrested in prophase I but resume meiosis upon ovulation. Competence to resume meiosis is not achieved until late in oocyte development and correlates with size, the appearance of a chromatin rim around the nucleolus and the ability to resume meiosis spontaneously when experimentally removed from the follicle^{13,14}. We evaluated meiotic competence of isolated oocytes by these criteria. The average diameter of $Cx37^{-/-}$ oocytes ($57 \pm 2 \mu$ m, $N = 21$ compared to $85 \pm 3 \mu$ m, $N = 19$ for controls) was below the size normally associated with meiotic competence¹³. In addition, although 87% ($N = 23$) of oocytes from wild-type antral follicles exhibited chromatin rimming, only 18% ($N = 33$) of those from $Cx37^{-/-}$ follicles were rimmed. Finally, 83% ($N = 12$) of wild-type oocytes but only 9% ($N = 34$) of $Cx37^{-/-}$ oocytes exhibited germinal vesicle breakdown, an easily scored indicator of meiotic resumption. Although the majority of $Cx37^{-/-}$ oocytes in the developmental stages we examined were clearly not competent, we have not ruled out the possibility that the oocytes slowly achieve competence

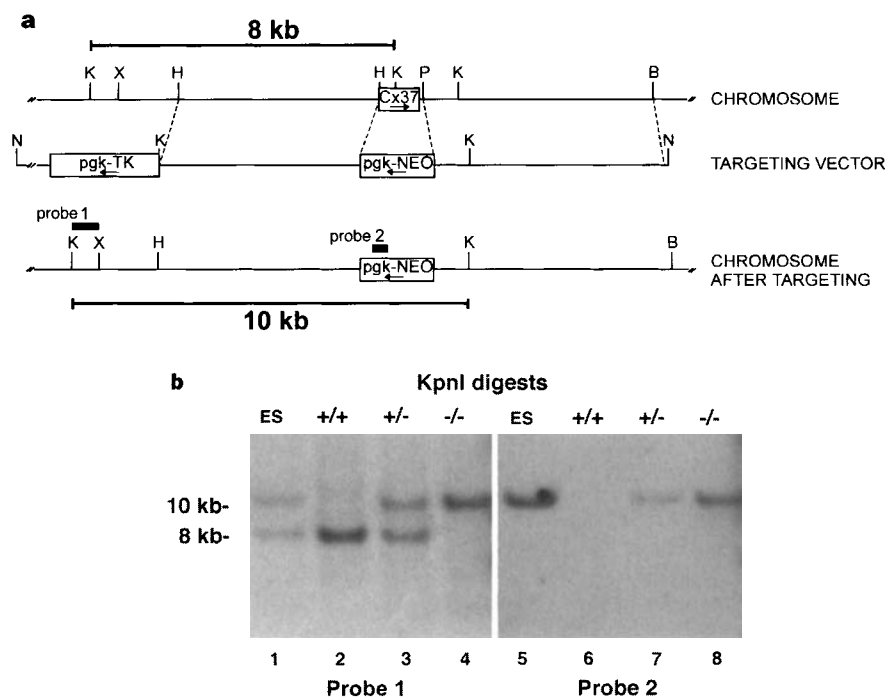


Figure 2 Targeting strategy for inactivation of the *Cx37* gene and Southern analysis of generated mice. **a**, The genomic structure near the *Cx37* locus. The *Cx37* coding region is contained within a single exon²⁴. Only relevant restriction enzyme sites are shown. A replacement targeting vector was designed to delete all of the coding region except for the first 77 bp. Homology regions are delimited by dashed lines. *KpnI* digestion of homologous recombinants yields a diagnostic 10 kb fragment that hybridizes with both probes, whereas wild-type DNA yields an

8 kb fragment that hybridizes only with probe 1. **b**, Homologously recombined ES cell clone DNA (lanes 1 and 5) and tail DNA from the generated mice (lanes 2–4 and 6–8) were digested with *KpnI* and analysed by Southern blotting using probes 1 and 2. The pattern of hybridization confirmed that correct targeting had occurred in the ES cell clone and that germ-line transmission was achieved. B, *BamHI*; H, *HindIII*; K, *KpnI*; N, *NotI*; P, *PmlI*; X, *XbaI*.

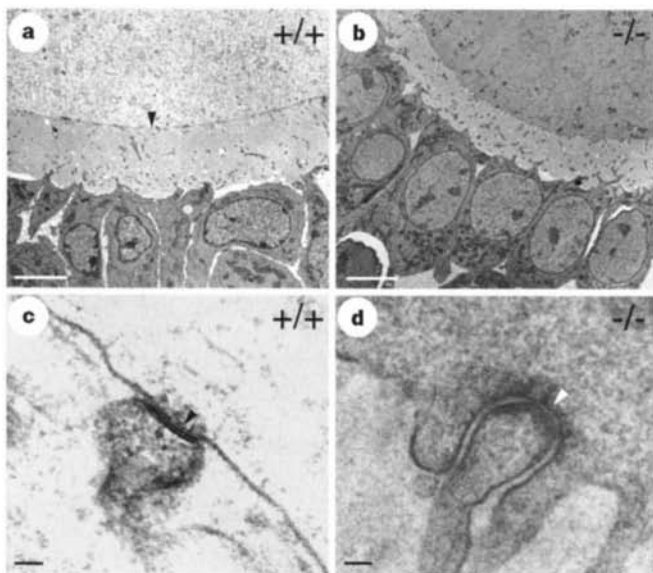


Figure 3 Electron microscopic analysis of ovarian follicles. **a**, Wild-type preantral follicle displaying (from top to bottom of panel) ooplasm, zona pellucida and granulosa cells. This orientation is maintained in all panels. Granulosa cells extend processes through the zona pellucida which contact the oolemma (arrowhead) and form both gap junctions and adherens junctions. Portions of the granulosa cell processes are visible within the zona pellucida as darkly staining areas. **b**, The zona pellucida and granulosa cell processes in *Cx37*^{-/-} follicles appeared grossly normal. **c**, At much higher magnification, gap junctions (arrowhead) between granulosa cell processes and oocytes were observed in wild-type follicles. **d**, *Cx37*^{-/-} follicles contained adherens junctions (arrowhead) between granulosa cells and the oocyte, but lacked gap junctions. Scale bars, 5 μ m (**a** and **b**); 50 nm (**c**); 70 nm (**d**).

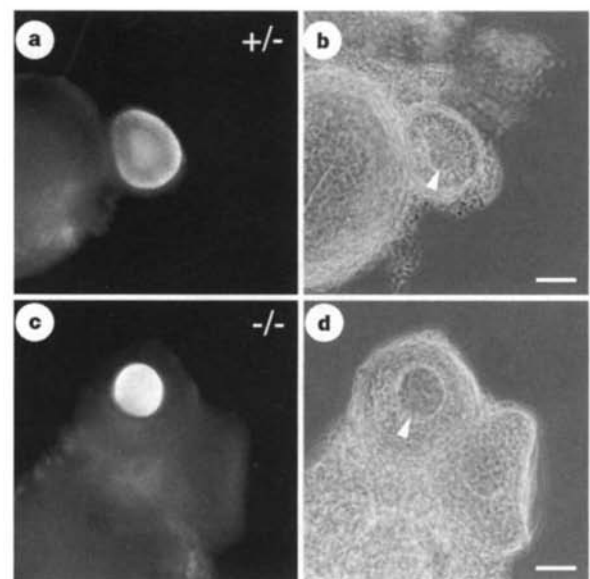


Figure 4 Loss of junctional coupling between oocyte and granulosa cells in *Cx37*^{-/-} mice. **a**, **b**, Neurobiotin was microinjected into a *Cx37*^{+/+} oocyte (arrowhead) of an isolated preantral follicle and then labelled with avidin-FITC. Neurobiotin was detected in both oocyte and surrounding granulosa cells, indicating junctional coupling. **c**, **d**, Neurobiotin microinjected into a *Cx37*^{-/-} oocyte (arrowhead) did not transfer to surrounding granulosa cells, indicating a lack of junctional coupling. Scale bars, 20 μ m (**a** to **d**).

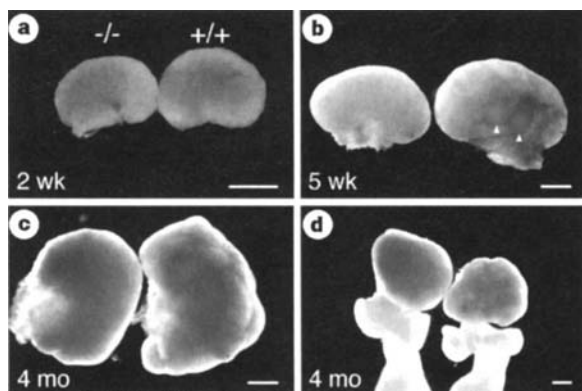


Figure 5 Anatomical differences between ovaries from *Cx37*^{-/-} and wild-type mice. In all panels, the ovary on the left is from a *Cx37*^{-/-} animal. **a**, Ovaries from 2-week-old *Cx37*^{-/-} and *Cx37*^{+/+} mice were similar in appearance. **b**, By 5 weeks, at sexual maturity, ovaries from wild-type mice had surface distensions (arrowheads) due to the presence of large follicles, whereas *Cx37*^{-/-} ovaries were smooth. **c**, At 4 months, *Cx37*^{-/-} ovaries were still abnormally smooth. **d**, No gross abnormalities were observed in oviducts or uterine horns of *Cx37*^{-/-} mice. Scale bars, 500 µm (**a** to **d**).

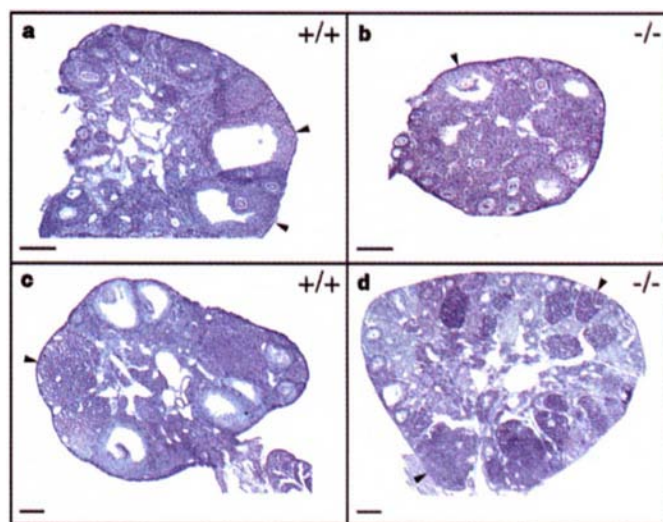


Figure 6 Histological appearance of ovaries from *Cx37*^{-/-} and wild-type mice. **a**, At 5 weeks, wild-type ovaries displayed several well-developed Graafian follicles (arrowheads). **b**, At 5 weeks, *Cx37*^{-/-} ovaries lacked Graafian follicles and only rarely displayed large antral follicles (arrowhead). **c**, At 4 months, wild-type ovaries exhibited several large corpora lutea (arrowhead), indicative of successful ovulation. **d**, At 4 months, *Cx37*^{-/-} ovaries still lacked Graafian follicles and ovulation was never detected. However, a large number of small corpus luteum-like structures (arrowheads) were evident. Scale bars, 200 µm (**a** to **d**).

and are then rapidly eliminated by an unknown mechanism and are thus not detected. Regardless, these results demonstrate that junctional communication strongly influences the acquisition of meiotic competence by the oocyte. Previous studies suggested that junctional communication maintains oocyte arrest in the first meiotic prophase by providing a mechanism for delivery of inhibitory signals^{4,15-17}. This model predicts that competent oocytes lacking junctional communication with granulosa cells would undergo premature meiotic resumption. Histological indications of premature resumption such as germinal vesicle breakdown or polar bodies were not observed (Fig. 6) but if full meiotic competence is not achieved by *Cx37*^{-/-} oocytes, then our data are unable to address the model.

Recent studies have linked connexin mutations to neurodegenerative disorders¹⁸ and cardiovascular defects^{19,20}. Although it remains to be determined whether defects in *Cx37* cause human infertility or other pathologies, *Cx37* distribution in human ovarian follicles is similar to mouse (data not shown). In addition, abnormalities observed in *Cx37*^{-/-} animals are similar to those exhibited in karyotypically normal spontaneous premature ovarian failure, a human disorder of unknown aetiology²¹. In particular, inappropriate luteinization is a specific feature of this disease, suggesting that alterations in connexin-mediated intercellular communication could be responsible. □

Methods

Immunocytochemistry and histology. Ovaries were fixed by perfusion of 1% formaldehyde through the left ventricle. Frozen sections (10 µm) were immunolabelled with anti-*Cx37* (ref. 9) and anti-*Cx43* (ref. 6) antibodies or stained with haematoxylin and eosin. Antigen-antibody complexes were detected with rhodamine-conjugated goat F(ab')₂ anti-rabbit IgG (Pierce).

Gene targeting and animal breeding. A rat *Cx37* complementary DNA (ref. 7) was used to isolate genomic clones from a 129/Sv mouse genomic library (Stratagene). A replacement targeting vector containing a *pgk-neo* cassette was designed to delete all of the coding region except for the first 77 base pairs. The vector contained 5' and 3' homology regions of 5 kilobases (kb) and 6 kb,

respectively, inserted into the pPNT vector²² and included a *pgk-TK* cassette upstream of the 5' homology region. The vector was electroporated into J1 ES cells and *neo*⁻-FIAU^r-clones were isolated by positive/negative selection as described²³. Homologous recombinants were identified by Southern blotting using a 5' probe (probe 1) and a probe within the *neo* gene (probe 2) and were obtained at a frequency of 1 in 20 *neo*⁻-FIAU^r colonies. The deleted allele was maintained in a mixed strain background (C57BL/6-129/Sv). However, strain dependence of the phenotype was evaluated after mating male chimaeras with pure 129/Sv females. All homozygous females tested were infertile regardless of strain background.

Electron microscopy. Ovaries were fixed by perfusion with 4% glutaraldehyde, 0.5% formaldehyde, 0.1 M cacodylate buffer (pH 7.3) and then fixed by immersion for 2 h at room temperature. Samples were further treated with osmium tetroxide, tannic acid, uranyl acetate and lead citrate before examination on a JEOL100 CX electron microscope.

Dye transfer. Ovaries were placed in Waymouth's medium MB752/1 (Gibco BRL) containing 0.1 mg ml⁻¹ collagenase and dissected apart with 30-gauge syringe needles. Isolated groups of preantral follicles were transferred to polylysine-coated tissue culture dishes containing medium and allowed to attach. Oocytes with stable resting potentials were iontophoretically microinjected with a 5% solution of neurobiotin (Vector Laboratories) for 5 min, with 5-nA pulses delivered every second on a 50% duty cycle. Following microinjection, the follicles were incubated for 30 min to allow for junctional transfer of neurobiotin then fixed in 4% formaldehyde. Neurobiotin was detected by staining with 0.5 µg ml⁻¹ avidin-FITC (Boehringer Mannheim) in 1 × PBS, 0.25% TX100, 2% BSA.

Evaluation of meiotic competence. Oocytes were released from antral follicles by dissection with 30-gauge syringe needles. Attached cumulus cells, if present, were removed by passing the oocyte through a drawn-out Pasteur pipette. Oocyte diameter measurements were performed using the Northern Lights computer-aided video micrography system. Chromatin rimming assays were performed by staining oocytes with 10 µg ml⁻¹ Hoechst 33258 dye (Molecular Probes) for 10 min as described¹⁴. Germinal vesicle breakdown assays were performed by incubating isolated oocytes in Waymouth's medium for 5–16 h and then scoring for the presence of intact vesicle by phase-contrast microscopy.

Received 13 September; accepted 10 December 1996.

1. Eppig, J. J. *Bioessays* **13**, 569–574 (1991).
2. Wassermann, P. M. & Albertini, D. F. in *The Physiology of Reproduction* (eds Knobil, E. & Neill, J. D.) 79–122 (Raven, New York, 1994).
3. Anderson, E. & Albertini, D. F. *J. Cell Biol.* **71**, 680–686 (1976).
4. Gilula, N. B., Epstein, M. L. & Beers, W. H. *J. Cell Biol.* **78**, 58–75 (1978).
5. Goodenough, D. A., Goliger, J. A. & Paul, D. L. *Annu. Rev. Biochem.* **65**, 475–502 (1996).
6. Beyer, E. C., Kistler, J., Paul, D. L. & Goodenough, D. A. *J. Cell Biol.* **108**, 596–605 (1989).
7. Haefliger, J. A. *et al. J. Biol. Chem.* **267**, 2057–2064 (1992).
8. Reed, K. E. *et al. J. Clin. Invest.* **91**, 997–1004 (1993).
9. Goliger, J. A. & Paul, D. L. *Dev. Dyn.* **200**, 1–13 (1994).
10. Valdimarsson, G., De Sousa, P. A. & Kidder, G. M. *Mol. Reprod. Dev.* **36**, 7–15 (1993).
11. Bjorkman, N. *Acta Anat.* **51**, 125–147 (1962).
12. El-Fouly, M. A., Cook, B., Nekola, M. & Nalbandov, A. V. *Endocrinology* **87**, 286–293 (1970).
13. Sorensen, R. A. & Wassarman, P. M. *Dev. Biol.* **50**, 531–536 (1976).
14. Mattson, B. A. & Albertini, D. F. *Mol. Reprod. Dev.* **25**, 374–383 (1990).
15. Cho, W. K., Stern, S. & Biggers, J. D. *J. Exp. Zool.* **187**, 383–386 (1974).
16. Dekel, N. & Beers, W. H. *Proc. Natl Acad. Sci. USA* **75**, 4369–4373 (1978).
17. Bornslaeger, E. A. & Schultz, R. M. *Biol. Reprod.* **33**, 698–704 (1985).
18. Bergoffen, J. *et al. Science* **262**, 2039–2042 (1993).
19. Britz-Cunningham, S. H., Shah, M. M., Zuppan, C. W. & Fletcher, W. H. *N. Engl. J. Med.* **332**, 1323–1329 (1995).
20. Reaume, A. G. *et al. Science* **267**, 1831–1834 (1995).
21. Nelson, L. M. *et al. J. Clin. Endocrinol. Metab.* **79**, 1470–1475 (1994).
22. Tybulewicz, V. L., Crawford, C. E., Jackson, P. K., Bronson, R. T. & Mulligan, R. C. *Cell* **65**, 1153–1163 (1991).
23. Li, E., Bestor, T. H. & Jaenisch, R. *Cell* **69**, 915–926 (1992).
24. Willecke, K. *et al. J. Cell Biol.* **114**, 1049–1057 (1991).

Acknowledgements. We thank J. Gabriels for affinity-purified preparations of anti-Cx37 and anti-Cx43 antibodies, S. Fields for advice about karyotyping ES cell clones, B. Quade for providing human ovarian tissue, D. Albertini for helpful discussions and M. Thompson for a critical reading of the manuscript. This work was supported by grants from the NIH (D.L.P., D.A.G., A.M.S.), Bristol Myer-Squibb (E.L.) and the American Heart Association (A.M.S.).

Correspondence and requests for materials to D.L.P. (e-mail: dpaul@warren.med.harvard.edu).

Spatio-temporal frequency domains and their relation to cytochrome oxidase staining in cat visual cortex

Doron Shoham^{*}, Mark Hübener[†], Silke Schulze[†],
Amiram Grinvald^{*} & Tobias Bonhoeffer[†]

^{*} Weizmann Institute of Science, Rehovot 76100, Israel

[†] Max-Planck Institute for Psychiatry, Am Klopferplatz 18A,
82152 München-Martinsried, Germany

Spatial and temporal frequencies are important attributes of the visual scene. It is a long-standing question whether these attributes are represented in a spatially organized way in cat primary visual cortex^{1–4}. Using optical imaging of intrinsic signals^{5–10}, we show here that grating stimuli of different spatial frequencies drifting at various speeds produce distinct activity patterns. Rather than observing a map of continuously changing spatial frequency preference across the cortical surface, we found only two distinct sets of domains, one preferring low spatial frequency and high speed, and the other high spatial frequency and low speed. We compared the arrangement of these spatio-temporal frequency domains with the cytochrome oxidase staining pattern, which, based on work in primate striate cortex, is thought to reflect the partition of the visual cortex into different processing streams. We found that the cytochrome oxidase blobs in cat striate cortex coincide with domains engaged in the processing of low spatial and high temporal frequency contents of the visual scene. Together with other recent results¹¹, our data suggest that spatio-temporal frequency domains are a manifestation of parallel streams in cat visual cortex, with distinct patterns of thalamic inputs and extrastriate projections.

Previous studies have suggested that the primary visual cortex performs a two-dimensional Fourier-like analysis of the retinal

image into its spatial frequency components^{1–4,12–18}. Single-unit recordings addressing the question of a clustering of cortical cells tuned to similar spatial scales came to somewhat contradictory results: while some studies reported a grouping of cells with similar spatial frequency tuning in a laminar fashion^{1,3}, others observed clustering but without any obvious relation to either cortical layers or columns⁴. On the other hand, 2-deoxyglucose (2-DG) mapping demonstrated a columnar activation pattern in response to a single spatial frequency stimulus². These results were interpreted as evidence for the existence of a cortical representation of spatial frequency resembling the cortical orientation map^{2,18}. This hypothesis was not tested with 2-DG because the responses to different frequencies could not be imaged in the same cortex. Moreover, from the 2-DG studies it was difficult to rule out the possibility that the 'spatial frequency map' was related to the ocular dominance map and caused chiefly by difference in acuity between the two eyes. Optical imaging of intrinsic signals in cat area 17 (Fig. 1a) enabled us to address these issues directly. Using sets of moving gratings of several orientations at two different spatial frequencies, we obtained the well-known orientation maps^{8,19} (data not shown). At each orientation the functional maps were similar but not identical for the two spatial frequencies. Consequently, proper analysis of the data showed the detailed organization of neuronal populations preferring low and high spatial frequencies and revealed a patchy activation pattern (Fig. 1b). A distinct segregation of light and dark regions is visible in this functional map. The dark regions are activated more strongly by the low spatial frequency, and the light regions are preferentially activated by the high spatial frequency. The amplitude of the spatial frequency maps was always low compared with the corresponding orientation preference maps; the average amplitude ratio of orientation maps to spatial frequency maps was 2.8 ± 0.3 (mean $\pm$ s.d., $n = 13$ hemispheres). Nevertheless, all these maps were highly reproducible (Fig. 1 legend). We also confirmed that they could not be attributed to a difference in visual acuity of the eyes as they were clearly different from the ocular dominance pattern (data not shown).

The layout of the spatial frequency domains varied: some of the domains were patch-like, whereas others had the shape of elongated bands. The distance between the domains representing similar spatial frequencies was roughly equal to the distance between iso-orientation domains in the same region of cortex. In many hemispheres the low spatial frequency domains tended to form isolated patches ('islands') embedded in a contiguous matrix ('sea') of high spatial frequency preference.

To confirm the maps obtained with optical imaging, we targeted electrophysiological recordings to the centres of the spatial frequency domains (Fig. 1c–e). We restricted them to the superficial layers and performed multi-unit rather than single-unit recordings because such data should be closest to optical imaging data, which also measure population activity in the supragranular layers. As expected from the reported broad distribution of optimal spatial frequencies of single cells at one cortical locus^{17,20}, the tuning curves of multi-unit activity were quite wide (~ 3 octaves, full-width at half height). We found that the tuning curves were nearly identical at different recording depths (150–1,000 μ m) of the same penetration (Fig. 1d), reinforcing the notion of a columnar arrangement of spatial frequency preferences (as opposed to laminar), at least in the superficial layers. When examining the tuning curves at the different penetrations (Fig. 1e) we found, despite the large widths of the tuning curves, a significant difference between the 'low' and the 'high' sites, confirming the optical imaging data. The widths of the multi-unit tuning curves, together with the large overlap between the curves of the 'low' and 'high' sites, may explain why previous unguided single-unit studies^{1,3,4} failed to reveal the columnar organization for spatial frequency. Moreover, it also explains why the amplitude of optically imaged spatial frequency maps is relatively low.

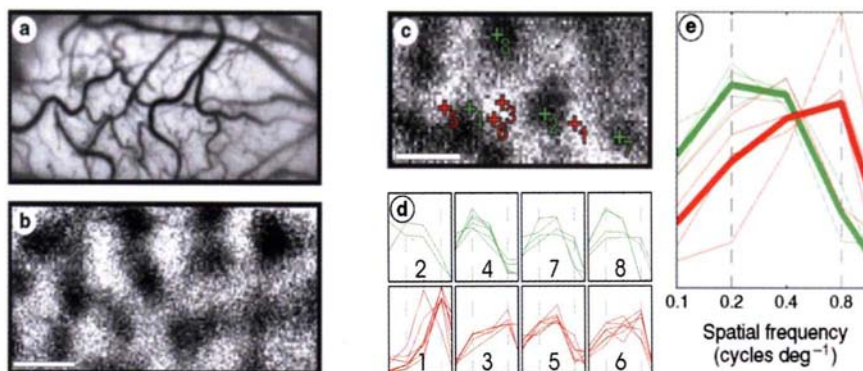


Figure 1 Spatial frequency domains in cat area 17. **a**, Image of the cortical surface (area 17, right hemisphere) obtained with green light (540 nm) to emphasize the blood vessel pattern. **b**, Optical imaging of spatial frequency domains. The differential activation map was produced by dividing the average image obtained with four drifting gratings of 0.2 cycles per deg (orientations of 0°, 45°, 90° and 135°) by the corresponding image from 0.6 cycles per deg stimuli. The temporal frequency was 2 Hz for all stimuli (similar results were obtained when the velocity was kept constant, see Fig. 3). The dark patches in the map correspond to regions that were activated more strongly by the low spatial frequency; the white areas were activated more by the high spatial frequency. The full grey scale corresponds to a fractional change of 2.4×10^{-4} . In comparison, the amplitude of the differential orientation maps computed from these data was ~ 3 times higher. Similar results were obtained in 13 hemispheres. To test the reproducibility of

these maps, we divided each data set into two halves (independent trials) and computed the maps separately from each set. We then compared the resulting images and verified that the layout of the functional domains was the same in both maps. The functional map in **a** is the average of 96 trials in which each stimulus condition was shown once. **c**, Spatial frequency map obtained from a different animal in the same way as the map in **b**. Dark regions again represent regions preferentially activated by low spatial frequencies. Electrode penetrations aimed at low spatial frequency domains are marked with green crosses, those in high spatial frequency domains with red crosses. **d**, The normalized tuning curves from all multi-unit recordings are shown (green for 'low' sites, red for 'high' sites). **e**, Average tuning curves from all penetrations (thin lines, colours as in **d**), and the averages for all 'low' and all 'high' penetrations (thick lines). Scale bars, 1 mm.

In the experiments described above we used only two spatial frequencies, making it impossible to decide whether there is a continuous representation of a whole range of spatial frequencies, as is the case for orientation preference, or whether these maps reflect a segregation into two distinct systems with different preferred spatial frequencies. To answer this question we stimulated with four different spatial frequencies, spanning a range of at least three octaves. The resulting maps from two such experiments, showing the activation pattern evoked by each of the four stimulus sets, are shown in Fig. 2: dark regions show stronger activation by a single spatial frequency compared with activation by all stimuli. Instead of the multiple sets of patches representing the entire scale of spatial frequencies, which would be expected for a continuous cortical map representing spatial frequency, we found a segregation into just two sets of domains. A comparison of the activity maps from the first experiment (Fig. 2a–d) shows that the two maps obtained in response to the low spatial frequencies (Fig. 2a, b) are very similar to each other, as are the two maps obtained in response to the high spatial frequencies (Fig. 2c, d). Moreover, the two pairs of maps are almost complementary to each other (the green crosses mark dark patches in the top two maps and white patches in the bottom two, and vice versa for the red crosses). The maps obtained from a second cat (Fig. 2e–h) show a similar result, but here we chose an intermediate frequency not used in the first experiment (Fig. 2g). At this frequency, a nearly uniform activity pattern was obtained, indicating that this particular spatial frequency activated the high and the low spatial frequency domains almost equally. We found no stimulus that preferentially activated an additional (third) set of domains as would be expected from a continuous representation. Note, however, that somewhat counterintuitively this result does not exclude the possibility that the optimal spatial frequency for populations of neurons changes smoothly across the cortex, like in a true spatial frequency map. It is theoretically possible that the loci that are the most active when a stimulus is presented are not necessarily those that prefer that stimulus over all others²¹.

Having found two sets of spatial frequency domains in cat visual

cortex, we investigated whether temporal frequency and velocity of stimuli are also represented in a similar fashion. To test this we used stimuli of two spatial frequencies, each moving at one of two temporal frequencies. We computed the differential activation maps for each of the six possible pairs (Fig. 3). All functional maps (except for Fig. 3f, which is nearly flat) show similar activity patterns, differing primarily in their amplitudes. This result indicates that spatial frequency is not the only variable distinguishing the two sets of domains: Fig. 3e shows that stimuli of the same spatial frequency but different temporal frequencies (and therefore speeds) produce a weak but detectable map that is very similar to the functional maps in which the spatial frequency varied (Fig. 3a–d). Moreover, the same pattern was obtained when each of the three parameters were kept constant (temporal frequency in Fig. 3a, b; velocity in Fig. 3d; spatial frequency in Fig. 3e), indicating that this map depends on more than just one of these parameters. Because for drifting gratings the values of any two of the three variables determine the third, the map actually represents a combination of all three variables. The maps (most clearly Fig. 3e) also show that the low spatial frequency patches were preferentially activated by the faster moving stimuli, and the high spatial frequency regions by the slow stimuli.

An orderly organization for spatial frequency preference has also been observed in macaque striate cortex, with neurons in the cytochrome oxidase blobs preferring lower spatial frequencies than neurons in the interblob regions^{22–25}. Cytochrome oxidase blobs have also been detected in cat primary visual cortex^{26–28}, although no functional correlate of the blobs has been found in this species so far. We therefore investigated whether the same relationship between cytochrome oxidase staining pattern and spatial frequency preference found in monkey striate cortex might also hold true in cat area 17. In seven of the experiments, post-mortem, we sectioned the fixed cortex and reacted it for cytochrome oxidase after obtaining the spatial frequency maps. Figure 4a shows a stained section cut horizontally through the visual cortex. Because cytochrome oxidase staining is generally weaker in cat than in monkey visual cortex^{26–28} we generated a contrast-enhanced image (Fig. 4b).

Figure 2 Two sets of spatial frequency domains. Functional maps obtained by stimulating with multiple spatial frequencies in two animals. Maps were calculated by dividing the average image obtained with drifting gratings of each of four spatial frequencies by the average image from all stimuli. The spatial frequencies were 0.1, 0.2, 0.6 and 1 cycles per degree in the first cat (**a–d**), and 0.1, 0.2, 0.4, 0.8 cycles per degree in the second cat (**e–h**). Four orientations were used for each set; the temporal frequency was 2 Hz for all stimuli. Dark regions in each map correspond to higher activation by the corresponding spatial frequency compared to all four spatial frequencies. To aid visual comparison between the maps, crosses were put at corresponding positions in all images from the same experiment. Green crosses are in low spatial frequency patches, red crosses in high spatial frequency patches. In the first experiment, similarity is high between the two low (**a, b**), as well as between the two high spatial frequency maps (**c, d**). The two low spatial frequency maps are almost complementary to the two high spatial frequency maps. To quantify these observations, we computed the 6 pairwise correlation coefficients between the four maps after smoothing them with a gaussian filter ($\sigma = 50 \mu\text{m}$). The average correlation within pairs (**a–b, c–d**) was 0.45, and the average correlation between pairs was -0.72 . A similar result is evident in the second experiment, where the two low spatial frequency maps (**e, f**) are similar to each other (correlation coefficient 0.61) and are complementary to the high spatial frequency map (**h**, average correlation -0.83). These numbers have to be compared with the highest value that can be obtained for identical maps under the same experimental conditions (the correlation between two maps produced by the same stimuli on two independent sets of trials); this value was 0.86. The map obtained with an intermediate spatial frequency (**g**, 0.4 cycles per degree) is nearly flat, apart from some residual blood vessel artefacts (average correlation to **e, f**: -0.44 ; correlation to **h**: 0.13). The maps are an average of 85 trials (**a–d**) or 96 trials (**e–h**). Scale bars, 2 mm.

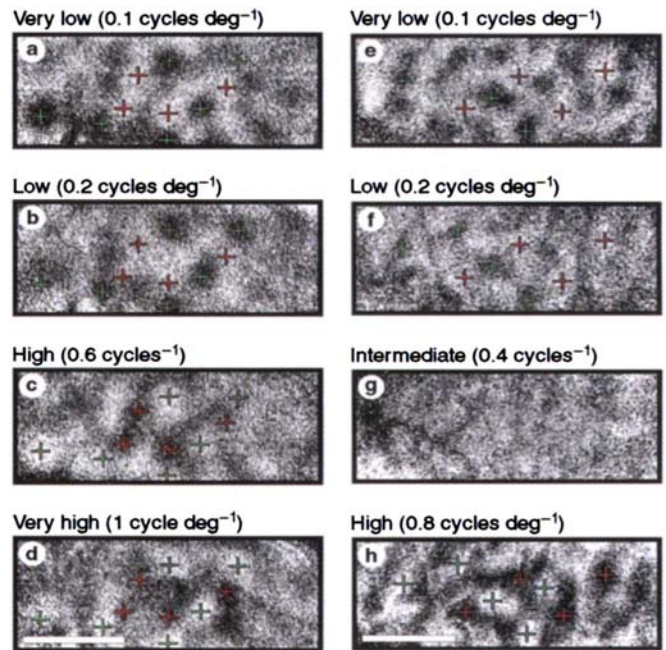
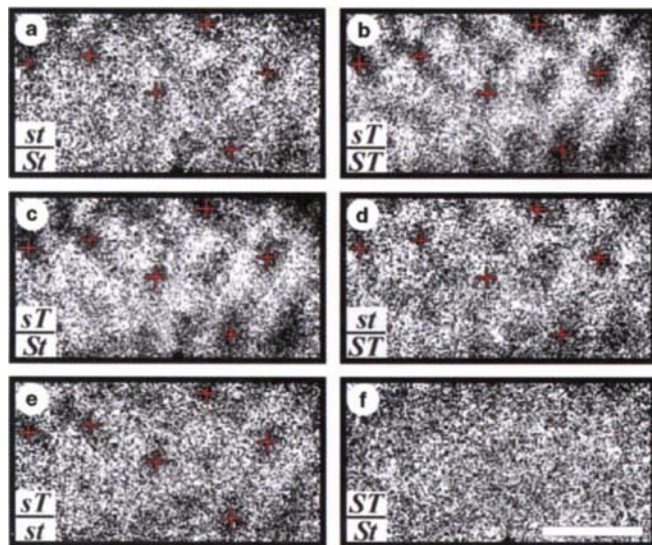


Figure 3 Spatio-temporal frequency domains. Maps produced by stimulation with two spatial frequencies, each moving at one of two temporal frequencies. The low spatial frequency (0.2 cycles per degree) is denoted by *s*, the high spatial frequency (0.6 cycles per degree) is denoted by *S*. The low and high temporal frequencies (0.66 Hz, 2 Hz) are denoted by *t* and *T*, respectively. The maps produced by dividing the images from the six possible pairs are shown. In the first four maps (**a–d**) the stimulus pairs had different spatial frequencies, with constant temporal frequency in **a, b** and constant velocity in **d**. In the last two maps (**e, f**) the spatial frequency was kept constant, and the temporal frequency and speed differed. In all the maps the low spatial frequency images were divided by the high frequency ones, and if the spatial frequency was the same (**e, f**), the images from the fast moving stimuli were divided by those from the slow moving stimuli. To aid visual comparison between the maps, crosses were put at the same positions in all images. The first five maps are very similar in shape and the last one is nearly flat. This was quantified using correlation coefficients as in Fig. 2. The average correlation coefficient between the map in **b**, which was taken as a reference, and those in **a, c–e** was 0.86; the correlation of **b** with **f** was -0.05 . Maps are averages of 64 trials. Scale bar, 2 mm.



This shows the cytochrome oxidase blobs as oval-shaped patches or bands of darker staining surrounded by regions of less intense staining. The spatial frequency map from the same cortical region is shown in Fig. 4c. Comparison of the histological section with the functional map shows that the cytochrome oxidase staining pattern coincides with the spatial frequency map: the blobs correspond to regions preferring low spatial frequencies and the lighter stained interblob regions correspond to high-spatial-frequency domains. This correspondence was found in all seven experiments. A second example is shown in Fig. 4d, e. Although the correspondence is not as obvious as in Fig. 4b, c, there is a close similarity between the two patterns.

Our results show that neurons in cat primary visual cortex cluster according to their spatial and temporal frequency preferences. These spatio-temporal frequency domains are in register with the

cytochrome oxidase staining pattern in cat primary visual cortex. The two neuronal populations may also differ in other functional aspects, for example in their contrast sensitivities. Thus, in cats, as in primates, neurons within cytochrome oxidase blobs are characterized by distinct combinations of receptive field properties.

Tracing studies have shown that blobs in cat visual cortex are specifically connected to other visual cortical areas^{29,30} and that they receive a strong input from—in the lateral geniculate nucleus¹¹. We have preliminary evidence that low spatial frequency domains also receive a stronger input from geniculate Y-cells³¹, which would be consistent with the preference for low spatial and high temporal frequencies revealed here. The blob and interblob regions of cat visual cortex, like their counterparts in primate visual cortex, appear to be compartments of parallel pathways that are specialized for analysing different attributes of the visual scene. □

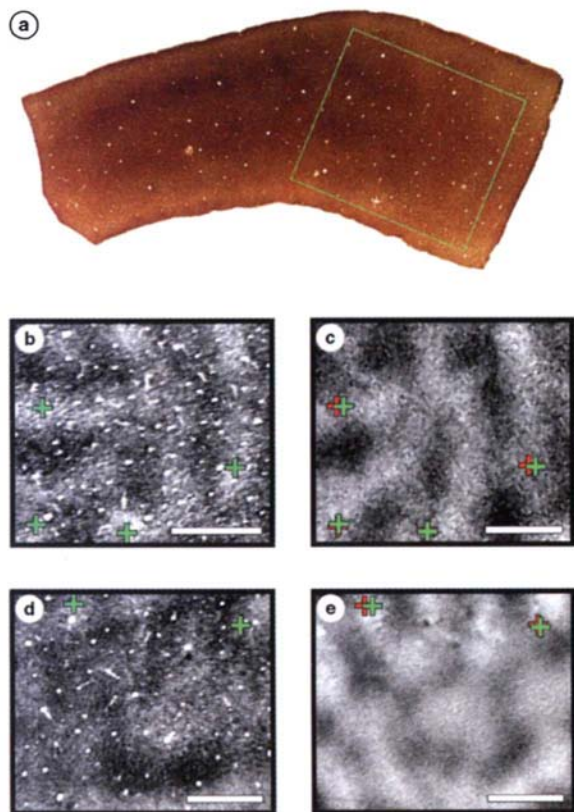


Figure 4 Correspondence between cytochrome oxidase staining pattern and spatio-temporal frequency domains in cat visual cortex. **a**, Cytochrome oxidase stained section cut horizontally through area 17 (no flattening of the tissue was done). In the right part of this image the section level passes through upper cortical layer 4. **b**, Contrast-enhanced image of a small portion of **a** (outlined in **a**). The cytochrome oxidase blobs are discernible as elongated patches or short bands of darker staining surrounded by regions of lower enzyme activity. The green crosses mark the injection holes resulting from small dye injections at the end of the optical imaging experiment to allow for the alignment of the functional maps with the stained histological sections. **c**, Spatial frequency map for the same region of visual cortex. Dark patches correspond to regions activated preferentially by low spatial frequencies (0.2 cycles per deg); light regions preferred high (0.6 cycles per deg) spatial frequency stimuli. Red crosses denote positions of dye injections into the cortex as judged from the *in vivo* blood vessel pattern. Green crosses mark injection holes recovered from histological sections (copied from **b**). Comparison of **b** and **c** shows that in most instances the blobs are in precise register with the low spatial frequency domains. **d**, **e**, Second example of the correlation between cytochrome oxidase staining (**d**) and spatial frequency map (**e**). Although the similarity is not as obvious as in **b** and **c**, close scrutiny reveals, for instance, a large inverted question mark slightly to the right of the middle of the images which is clearly visible in both pictures. Scale bars, 1 mm.

Methods

The preparation and maintenance of animals and optical imaging procedures were essentially as described^{7,8,10}. Experiments were done on 16 kittens (age 6–10 weeks) and 3 adult cats.

Surgery. Animals were initially anaesthetized with ketamine (15–30 mg per kg, i.m.) and xylazine (1.5–2 mg per kg, i.m.), supplemented by atropine (0.15 mg per kg, i.m.). Sustained anaesthesia was achieved either by continuous intravenous infusion of sodium pentothal (4 mg kg⁻¹ h⁻¹) or by ventilating the animal with a mixture of N₂O, O₂ and halothane (40–50% N₂O, 50–60% O₂, 0.8–1.5% halothane). Animals were paralysed by intravenous infusion of flaxedil (30 mg kg⁻¹ h⁻¹) or succinylcholine hydrochloride (20 mg kg⁻¹ h⁻¹). ECG, EEG, arterial oxygen saturation and rectal temperature were monitored

during the experiment; expired CO₂ was kept at ~3.5%. The eyes were refracted carefully and corrected with appropriate contact lenses. The skull was opened above area 17 and the part corresponding to an eccentricity of 0° to ~5° was exposed for imaging. The dura was removed and a stainless steel chamber was cemented onto the skull. The chamber was sealed with wax, filled with silicon oil and closed with a coverglass.

Visual stimuli. Large field stimuli consisted of moving high-contrast, square-wave or sine-wave gratings presented binocularly or monocularly at four orientations and different spatial and temporal frequencies in a pseudorandom sequence. Each stimulus was repeated 64 to 160 times.

Optical imaging. Cortical maps were visualized using optical imaging based on intrinsic signals as described elsewhere¹⁰. In short, the cortex was illuminated with red light (630 or 707 nm) and cortical images were acquired while the animal was visually stimulated, using either a slow-scan CCD-camera or a differential video imaging system (ORA-2001 or Imager-2001; Optical Imaging).

Computation of maps. Spatial frequency maps were calculated by adding all orientation maps of one spatial frequency and dividing the result by the sum of the orientation maps of the second spatial frequency. To obtain the spatial frequency maps we usually presented the stimuli binocularly. In most experiments we included a part in which the same stimuli were applied to each of the eyes separately. This was done to rule out the possibility that poor refraction in one eye might impair the response of that eye to high spatial frequencies, in which case the 'spatial frequency map' would actually reflect the ocular dominance pattern. To exclude this possibility we each computed the spatial frequency maps for eye separately and verified that each map was similar to the binocular map. We also used the monocular data to compute the ocular dominance map, which in all cases was very different from the spatial frequency map. In some cases, functional maps were high-pass filtered with a gaussian filter ($\sigma = 280 \mu\text{m}$) to remove haemodynamic noise.

Electrophysiology. We used a chamber designed to allow fine, three-dimensional manipulation of an electrode within the chamber (A. Arieli and A.G., manuscript in preparation). Multi-unit spikes were collected using a window discriminator. In each penetration, between 2 and 10 recordings at depths between 150 and 1,000 μm depths were made. At the start of each recording we used a hand-held projector to determine the receptive field position and the optimal orientation for the recorded units. The screen was then centred on the receptive field and full-screen computer-generated gratings were used to determine the spatial frequency tuning curve. The gratings were presented at the optimal orientation at each of 5 spatial frequencies (0.1, 0.2, 0.4, 0.8, 1.2 cycles per deg, drifting at 2 Hz) in a pseudorandom order. All stimuli were presented monocularly to the same eye to avoid confounding ocular dominance effects due to a possible difference in refraction of the two eyes. Each tuning curve was normalized by dividing all spike counts by the average. To create an average tuning curve for each penetration site (Fig. 1e), normalized curves from all depths were averaged.

Cytochrome oxidase histochemistry. At the end of the optical imaging experiment, 4 to 6 injections of highly diluted red and green latex beads were placed in the cortex to provide landmarks for later alignment of the maps with the histological sections. After perfusion with saline and fixative (4% paraformaldehyde in 0.1 M phosphate buffer), frozen sections (60 μm) were cut parallel to the focus plane of the camera and reacted for cytochrome oxidase^{32,33}.

Received 7 August; accepted 6 December 1996.

- Maffei, L. & Fiorentini, A. *Vision Res.* **17**, 257–264 (1977).
- Tootell, R. B. H., Silverman, M. S. & De Valois, R. L. *Science* **214**, 813–815 (1981).
- Berardi, N., Bisti, S., Cattaneo, A., Fiorentini, A. & Maffei, L. *J. Physiol. (Lond.)* **323**, 603–618 (1982).
- Tolhurst, D. J. & Thompson, I. D. *Exp. Brain Res.* **48**, 217–227 (1982).
- Grinvald, A., Lieke, E., Frostig, R. D., Gilbert, C. D. & Wiesel, T. N. *Nature* **324**, 361–364 (1986).
- Frostig, R. D., Lieke, E. E., Ts'o, D. Y. & Grinvald, A. *Proc. Natl Acad. Sci. USA* **87**, 6082–6086 (1990).
- Ts'o, D. Y., Frostig, R. D., Lieke, E. E. & Grinvald, A. *Science* **249**, 417–420 (1990).
- Bonhoeffer, T. & Grinvald, A. *J. Neurosci.* **13**, 4157–4180 (1993).
- Malonek, D. & Grinvald, A. *Science* **272**, 551–554 (1996).
- Bonhoeffer, T. & Grinvald, A. in *Brain Mapping: The Methods* (eds Toga, A. W. & Mazziotta, J. C.) 55–97 (Academic, San Diego, 1996).
- Boyd, J. D. & Matsubara, J. A. *J. Comp. Neurol.* **365**, 659–682 (1996).
- Campbell, F. W., Cooper, G. F. & Enroth-Cugell, C. *J. Physiol. (Lond.)* **203**, 223–235 (1969).
- Pollack, D. A., Lee, J. R. & Taylor, J. H. *Science* **173**, 74–77 (1971).
- Glezer, V. D., Ivanoff, V. A. & Tscherbach, T. A. *Vision Res.* **13**, 1875–1904 (1973).
- Maffei, L. & Fiorentini, A. *Vision Res.* **13**, 1255–1267 (1973).
- Robson, J. G. in *Handbook of Perception* Vol. V (eds Carterette, E. C. & Friedman, M. P.) 81–116

- (Academic, New York, 1975).
17. Töhlhurst, D. J. & Thompson, I. D. *Proc. R. Soc. Lond. B* **213**, 183–199 (1981).
 18. De Valois, R. L. & De Valois, K. K. *Spatial Vision* (Oxford University Press, Oxford, 1988).
 19. Bonhoeffer, T. & Grünwald, A. *Nature* **353**, 429–431 (1991).
 20. Movshon, J. A., Thompson, I. D. & Töhlhurst, D. J. *J. Physiol. (Lond.)* **283**, 101–120 (1978).
 21. Shoham, D. *Israel J. Med. Sci. (Abstr.)* **32**, S7 (1996).
 22. Tootell, R. B. H., Silverman, M. S., Hamilton, S. L., Switkes, E. & De Valois, R. L. *J. Neurosci.* **8**, 1610–1624 (1988).
 23. Silverman, M. S., Grosof, D. H., De Valois, R. L. & Elfar, S. D. *Proc. Natl Acad. Sci. USA* **86**, 711–715 (1989).
 24. Born, R. T. & Tootell, R. B. H. *Proc. Natl Acad. Sci. USA* **88**, 7066–7070 (1991).
 25. Edwards, D. P., Purpura, K. P. & Kaplan, E. *Vision Res.* **35**, 1501–1523 (1995).
 26. Murphy, K. M., Van Sluyters, R. C. & Jones, D. G. *Soc. Neurosci. Abstr.* **16**, 292 (1990).
 27. Dyck, R. H. & Cynader, M. S. *Proc. Natl Acad. Sci. USA* **90**, 9066–9069 (1993).
 28. Murphy, K. M., Jones, D. G. & Van Sluyters, R. C. *J. Neurosci.* **15**, 4196–4208 (1995).
 29. Boyd, J. D. & Matsubara, J. A. *Soc. Neurosci. Abstr.* **18**, 298 (1992).
 30. Matsubara, J. A. & Boyd, J. D. *Soc. Neurosci. Abstr.* **20**, 1742 (1994).
 31. Hübener, M., Schulze, S. & Bonhoeffer, T. *Soc. Neurosci. Abstr.* **22**, 951 (1996).
 32. Wong-Riley, M. T. T. *Brain Res.* **171**, 11–28 (1979).
 33. Hübener, M. & Bolz, J. *J. Comp. Neurol.* **324**, 67–80 (1992).

Acknowledgements. D.S. and M.H. contributed equally to this work. We thank F. Brinkmann, G. Brändle, D. Ettner, S. Leytus and C. Wijnbergen for technical support, and R. Tootell for constructive comments on the manuscript. Supported by grants from the Israel Science Foundation (A.G.) and the Max Planck Gesellschaft (T.B.). Additional support came from the Mijan Foundation (A.G.), the Human Frontier Science Program (T.B.), and Ms Enoch (A.G.).

Correspondence and requests for materials should be addressed to M.H. (e-mail: mark@neuro.mpg.de).

Synaptic tagging and long-term potentiation

Uwe Frey & Richard G. M. Morris*

Federal Institute for Neurobiology, Gene Regulation and Plasticity, PO Box 1860, Brennekestrasse 6, 39008 Magdeburg, Germany

* Centre for Neuroscience, University of Edinburgh, Crichton Street, Edinburgh EH8 9LE, UK

Repeated stimulation of hippocampal neurons can induce an immediate and prolonged increase in synaptic strength that is called long-term potentiation (LTP)—the primary cellular model of memory in the mammalian brain¹. An early phase of LTP

(lasting less than three hours) can be dissociated from late-phase LTP by using inhibitors of transcription and translation^{2–8}. Because protein synthesis occurs mainly in the cell body^{9–12}, whereas LTP is input-specific, the question arises of how the synapse specificity of late LTP is achieved without elaborate intracellular protein trafficking. We propose that LTP initiates the creation of a short-lasting protein-synthesis-independent ‘synaptic tag’ at the potentiated synapse which sequesters the relevant protein(s) to establish late LTP. In support of this idea, we now show that weak tetanic stimulation, which ordinarily leads only to early LTP, or repeated tetanization in the presence of protein-synthesis inhibitors, each results in protein-synthesis-dependent late LTP, provided repeated tetanization has already been applied at another input to the same population of neurons. The synaptic tag decays in less than three hours. These findings indicate that the persistence of LTP depends not only on local events during its induction, but also on the prior activity of the neuron.

Suppose that induction of early or late LTP on one input (S1) causes the transient activation of a local synaptic tag, but only repeated tetanization leading to late LTP initiates protein synthesis. Provided the creation of the putative tag is independent of protein synthesis, and plasticity-related proteins activated by LTP travel nonspecifically in the neuron, the proteins sequestered by a tag do not need to have been induced by the same event that sets the synaptic tag (Fig. 1).

To explore this idea, two independent synaptic inputs to the same neuronal population were stimulated in the CA1 region of hippocampal slices *in vitro* (Fig. 2a). Repeated tetanization of S1 (ref. 13) resulted in the establishment of late LTP (Fig. 2b, open circles) which lasted for at least 8 h, with minimal effect upon the second control input (Fig. 1a; Fig. 2b, S2, squares). A separate control experiment showed that consecutive induction of late LTP to S1 and S2 (1-h interval) was prevented by inhibitors of protein synthesis applied from 25 min before tetanization of input S1 until 1 h after initiation of LTP in S2 (Fig. 2c). The potentiation of the field

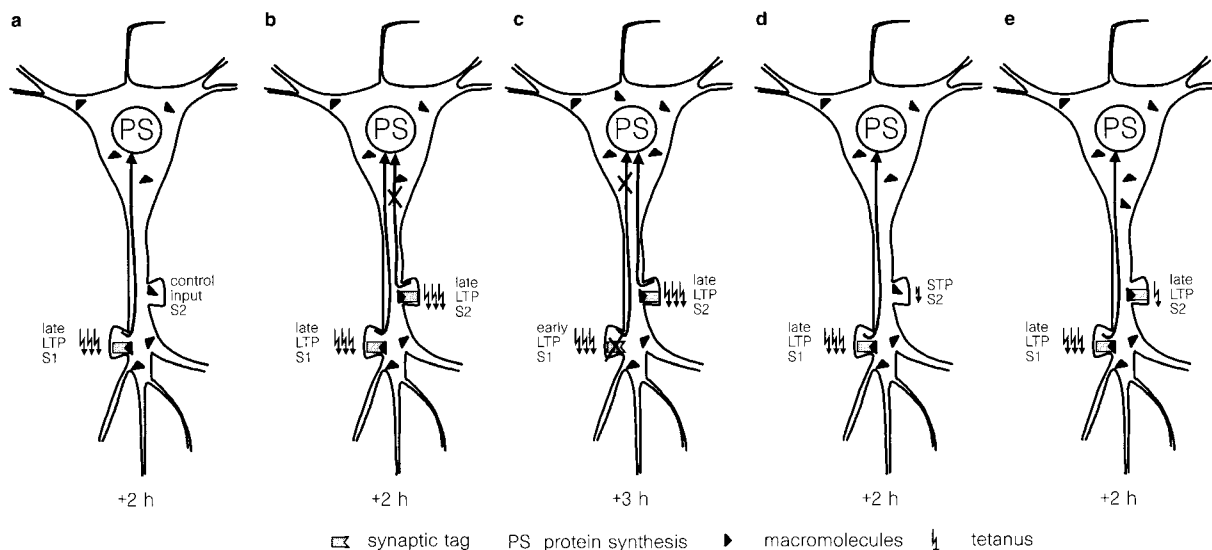


Figure 1 Possible occurrence of the synaptic tag (flag symbol) and proteins (triangles) after distinct stimulation of the two inputs at various time points after tetanization of synaptic inputs S1. **a**, Situation two hours after the induction of late LTP in S1. Intracellular arrows indicate the initiation of protein synthesis. Low-frequency control stimuli to S2 do not create a tag at this synaptic site. **b**, Late LTP is induced in S1 followed by application of a protein synthesis inhibitor that prevents the synthesis of macromolecules normally initiated by induction of late LTP in S2. Late LTP is nevertheless seen at S2 because its tag hijacks proteins synthesized by S1. The cross represents inhibition of protein synthesis. **c**,

Tetanization of S1 during protein synthesis inhibition transiently creates a tag that lasts for less than 3 h. Initiation of late LTP in S2 three hours after tetanization of S1 (and the removal of anisomycin) fails to rescue late LTP in S1. The tag created by tetanization of S1 must have decayed by the time the macromolecules associated with tetanization of S2 are available. **d, e**, Possible lack of a tag after STP-inducing stimulation and its association with early LTP after the induction of late LTP in S1. Very weak tetanization (small symbol in **d**) only leads to STP, whereas stronger single tetanization of S2 (**e**) results in late LTP.

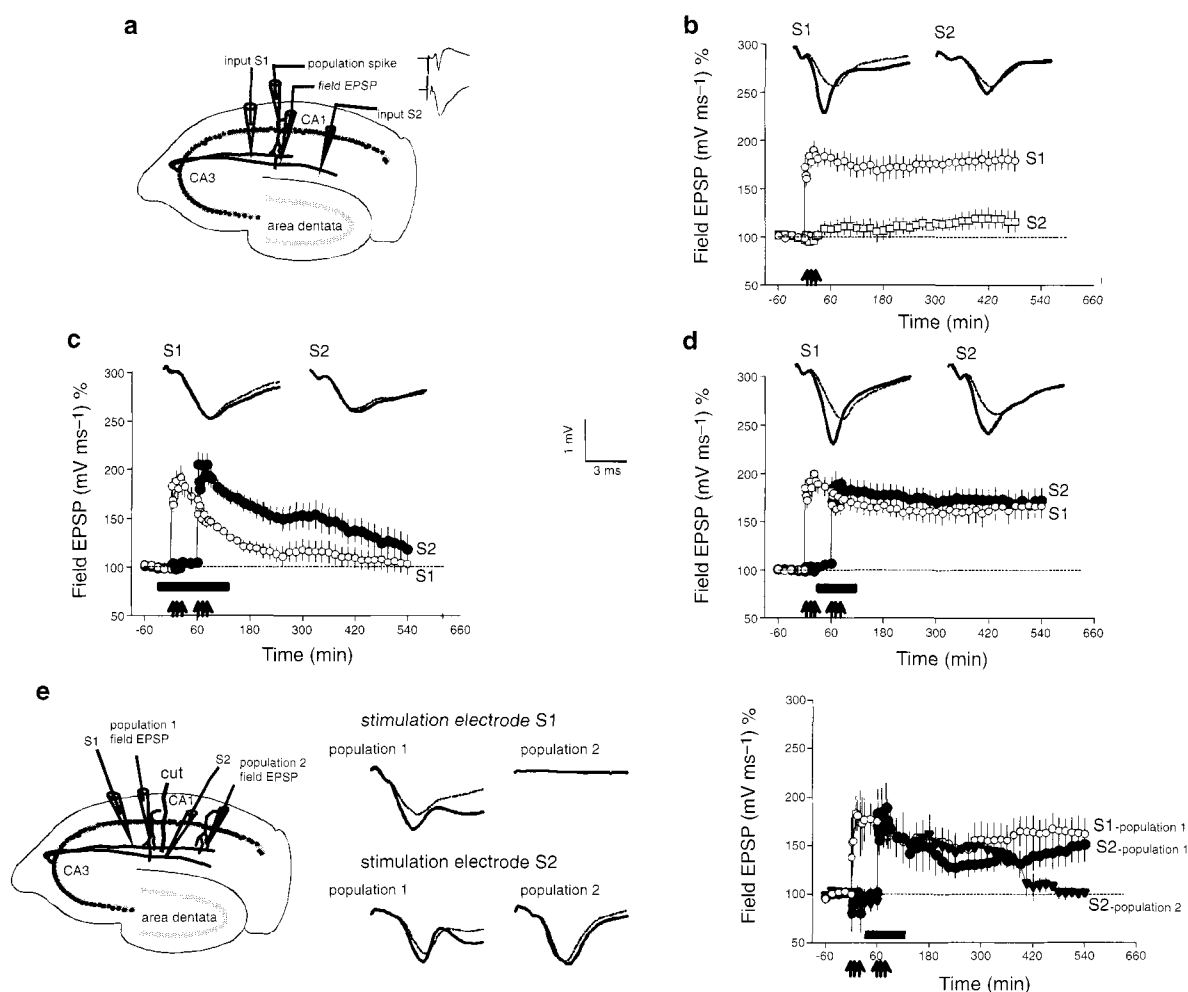


Figure 2 Induction of protein-synthesis-dependent LTP in the presence of a protein-synthesis inhibitor. **a**, Transversal hippocampal slice showing the positioning of the electrodes. The two independent inputs S1 and S2 to the same neuronal population, and the recording sites for the population spike amplitude and the field EPSP are shown. **b**, Tetanization of S1 (arrows) by a repeated-stimulation protocol revealed the establishment of late LTP (per cent change of field EPSP; open circles, 179.5 ± 10.7 and of the population spike amplitude $296.6 \pm 16.3\%$ of baseline after 8 h, $n = 6$; latter not shown). Low-frequency stimulation of S2 revealed stable recordings for the full 8 h after induction of LTP in S1 ($120.6 \pm 18.0\%$; $115.9 \pm 11.4\%$ respectively). Insets show representative recorded potentials immediately before (dotted line) and 7 h after tetanization (the same combination of potentials is shown for each experiment, calibration 1 mV per 3 ms). **c**, Induction of early LTP in two separate inputs after inhibition of protein synthesis revealed no late LTP in both measured parameters (field EPSP: in S1, $103.9 \pm 10.9\%$; S2, $118.1 \pm 14.3\%$ after 8 h, both statistically significantly different compared to control LTP in **b**; $n = 8$). **d**, LTP was induced in

S1 without drug application (open circles, $n = 7$). 35 min after tetanization of S1, anisomycin was added (bar), and 1 h after LTP of S1, input S2 was tetanized, but when protein synthesis was inhibited (filled circles). Paradoxically late LTP on S2 was still observed. **e**, Within-slice experiment showing the time course of LTP simultaneously in two different neuronal populations without or during protein synthesis inhibition. Left, positioning of the electrodes and the location of the microsurgical cut made to stimulate and record from two different neuronal populations 1 and 2. As for **d**, LTP was induced in S1 of population 1 (open circles, $n = 5$; field EPSP after 8 h $166.0 \pm 12.8\%$). Anisomycin (bar) was then applied and LTP was induced in S2 of population 1 and S1 of population 2. Late LTP in S2 of population 1 (filled circles; 8 h after tetanization: $145.2 \pm 15.9\%$) was rescued by LTP in S1 of population 1, whereas LTP in S2 of population 2 decayed as in **c** (filled triangles; $102.0 \pm 5.6\%$). The slightly different time course of LTP in S2 population 1 compared to **d** might be due to suboptimal stimulation parameters for this particular input (see Methods).

excitatory postsynaptic potential (EPSP) (and population spike; not shown) declined in both inputs as described previously³.

In the first key set of experiments, LTP in input S1 was induced (Fig. 2d, open circles) and anisomycin added to the bath medium 35 min later (a time when the protein-synthesis inhibitor is ineffective in influencing late LTP in this input³). Twenty-five minutes later, S2 was tetanized, but this time during the inhibition of protein synthesis (Fig. 2d, filled circles). The establishment of late LTP in S2 was normal, suggesting that proteins synthesized by LTP in S1 also allow the induction of late LTP in S2 (Fig. 1b). Similar results were obtained using a structurally different inhibitor of macromolecule synthesis (emetine; see Methods).

The paradoxical induction of protein-synthesis-dependent LTP during inhibition of protein synthesis was supported by a further within-slice experiment. Stimulating electrodes were positioned in stratum radiatum to activate simultaneously two independent inputs in one neuronal population and one input to a second, independent neuronal population (Fig. 2e). One electrode was positioned to evoke only potentials in S1 of population 1 (Fig. 2e, right panel, open circles) by means of a microsurgical cut through most of the Schaffer collaterals between the two populations recorded. The second electrode was positioned so that it simultaneously stimulated population 2 (Fig. 2e, filled triangles) and input S2 of population 1 (filled circles in Fig. 2e). Late LTP was first

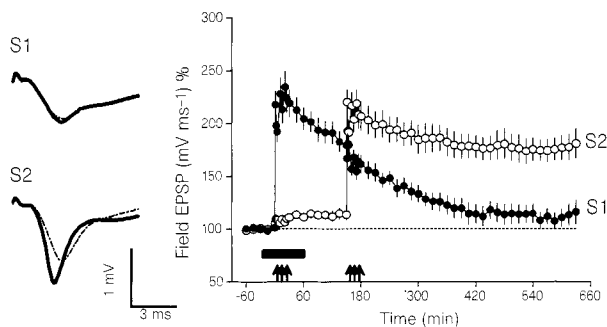


Figure 3 Transient occurrence of the putative synaptic tag. Experimental conditions were the same as for Fig. 2a. When anisomycin (bar) was applied 25 min before induction of LTP on S1 until 1 h afterwards (filled circles; $n = 6$), late LTP was prevented. After removal of the drug, tetanization of S2 (open circles) was made 2 h and 30 min after tetanization of S1. The persistence of late LTP on S2 was unaffected (field EPSP in S2: $166.2 \pm 12.3\%$ after 8 h). The tag created by tetanization of S1 must have decayed by the time the macromolecules associated with tetanization of S2 are available.

induced in input S1 of population 1 (Fig. 2e, open circles) and anisomycin applied 35 min later. Subsequent tetanization through stimulation electrode 2 resulted in late LTP in input S2 of the already activated neurons in population 1, but induced only a decaying early LTP in the independent population 2. This result strongly supports our proposition of a localized presence of proteins synthesized by LTP of input S1 in population 1 which could be hijacked by the proposed synaptic tag of input S2 to that population, but which are absent in the 'virgin' neuron population 2 because of the simultaneous inhibition of protein synthesis.

Is the putative synaptic tag permanently or transiently created after LTP induction? If transient, a time window should exist for the coincidence of the activated synaptic tag and newly synthesized proteins. To investigate this possibility, S1 was tetanized in the presence of anisomycin (Fig. 3, filled circles). The drug was then washed out for 2.5 h before S2 was tetanized (Fig. 3, open circles). Late LTP was induced in S2 and maintained for 8 h, but the LTP in S1 continued to decline to baseline. This finding suggests that proteins synthesized by tetanization of S2 about 3–4 h after LTP induction in S1 are unable to stabilize S1's potentiation. The putative synaptic tag is therefore transient, lasting no more than 2–3 h (Fig. 1c). This time window resembles the phase of LTP during which activated synapses are unable to respond with additional potentiation following a second tetanus¹⁴. Further tetanization can only induce additional potentiation during late LTP. The ability to react with further synaptic plasticity after prior induction of LTP may depend on the resetting of the proposed synaptic tag.

If the input specificity of late LTP is determined by the synaptic tag and its persistence by the availability of relevant proteins, the question arises whether short-term potentiation (STP) can be transformed into late LTP by prior induction of late LTP in another input to the same neuronal population. As late LTP normally requires repeated tetanization¹³, a weak single tetanus was used (see Methods) to induce a potentiation lasting for about 1–2 h (STP; Fig. 4a). Prior induction of late LTP in a first input S1 (Fig. 4b, open circles) did not prolong the potentiation of STP in S2 (Figs 4b, filled circles, and 1d). In the second protocol, a stronger but still single tetanus was applied to S2 that normally elicits early LTP lasting for 3–5 h (Fig. 4c) and corresponding to the stage named 'LTP1' in the model of ref. 1. It could be transformed into late LTP by prior induction of late LTP in another input (Figs 4d and 1e).

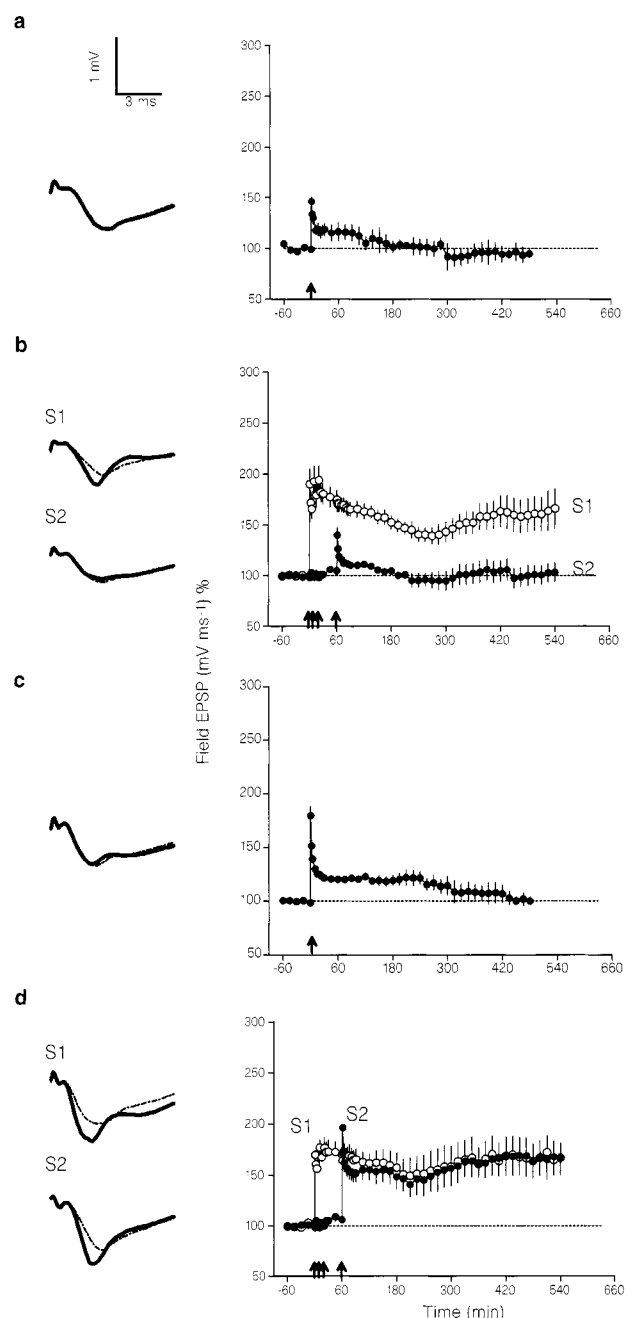


Figure 4 conversion of early-LTP into late-LTP. **a**, A weak tetanization protocol was used which elicited a short-term potentiation of the field EPSP for ~1 to 2 h ($n = 4$; symbols as in Fig. 1). **b**, Prior induction of late LTP in S1 (open circles; $n = 6$) did not influence the short potentiation in S2 (filled circles) induced 1 h after LTP on S1. **c**, A stronger single tetanization revealed early LTP lasting ~4 to 6 h ($n = 5$). **d**, When repeated tetanization and subsequent induction of late LTP in input S1 (open circles; $n = 5$) preceded the stronger single tetanization in S2 (filled circles), early LTP was transformed into late LTP in that input (field EPSP in S2: $167.7 \pm 17.6\%$ after 8 h).

To summarize, the consolidation of LTP involves the commitment of cellular proteins but this is insufficient to induce synapse-specific late LTP. A synaptic tag must also be transiently activated during early LTP. Our data provide the first experimental evidence for the selective catching of proteins by activated synapses, enabling late LTP to have input-specific properties without elaborate protein trafficking. They also establish that the putative synaptic tag

involves neither somatic nor local¹⁵ dendritic protein synthesis. Tetanization need not involve multiple trains, but must reach some critical threshold for the tag to be formed. Constitutively active kinases at activated synapses¹⁶ are appropriate candidates for fulfilling this function, particularly in the light of our findings that a tag lasts at most for 3 h and must be reset for further potentiation to occur. It remains to be determined whether the synaptic tag can also be reset in an activity-dependent way, for example by stimulation that induces depotentiation¹⁷.

These novel determinants of long-lasting potentiation have intriguing implications for the concepts of input specificity and persistence of LTP. According to the Hebbian model¹⁸, pre- and postsynaptic neurons have to be coactive within a distinct time to modify synaptic strength. Our results extend this hypothesis in one important respect: whereas the time interval for coactivity to create the synaptic tag must be short (<300 ms; refs 19, 20), the persistence of LTP can be influenced heterosynaptically by tetanization of other synaptic afferents over a period of 2–3 h. The establishment of late LTP appears to be determined by more than the local degree of NMDA-receptor activation²¹ in one synaptic input. The processing of incoming signals at a given time may alter the persistence of short-term changes in plasticity induced later at different synaptic inputs to the same neuron. Studies investigating the long-term modification of synaptic inputs should therefore take into account the prior and, possibly, the subsequent activity state²² of hippocampal neurons, considering them as integrative units over time²³. Processes other than potentiation-like events, such as long-term depression, might also influence different activated inputs in a similar way. These findings also suggest a new way in which protein synthesis can influence memory consolidation²⁴. A further, and speculative implication is that, if late-LTP-like synaptic changes are involved in long-term memory storage in the mammalian brain, our findings could provide a mechanistic explanation of why inconsequential events are typically remembered for much longer if they occur around the same time as well-remembered events. The vividness of 'flashbulb' memories²⁵ may derive, in part, from this cellular mechanism. □

Methods

53 transversal hippocampal slices (400 µm) were prepared from 53 male Wistar rats (7 weeks old) as described^{3,8,14}. Slices were incubated in an interface chamber at 32 °C and at a flow rate of artificial cerebrospinal fluid (ACSF, containing 124 mM NaCl, 4.9 mM KCl, 1.2 mM KH₂PO₄, 2.0 mM MgSO₄, 2.0 mM CaCl₂, 25.6 mM NaHCO₃, 10 mM D-glucose; carbogen consumption: 34 litres per hour) of 1 ml min⁻¹. In most of the experiments, two monopolar, lacquer-coated, stainless-steel electrodes (A-M Systems) were positioned in the stratum radiatum of the CA1 region for stimulation. For recording, 2 electrodes were placed in the CA1 dendritic and cell-body layer of a single neuronal population, except for the within-slice experiments (Fig. 2e), in which the two recording electrodes were positioned in the stratum radiatum at a distance that guaranteed—in combination with a microsurgical cut through a part of the CA1 region immediately after preparation—the stimulation of two independent neuronal populations. In the within-slice series, one stimulation electrode 1 was positioned only to evoke potentials in S1 of population 1. The second stimulation electrode simultaneously elicited potentials in S2 of population 1 and S1 of population 2 (Fig. 2e). In all other experiments, 2 stimulation electrodes were positioned at an adequate distance to stimulate separate inputs of one neuronal population (Fig. 2a). Slices were preincubated for at least 4–5 h, a period that is critical for a stable long-term recording for at least 8 h. This may be due to the requirement for a distinct stable level of activity to be reached after nonspecific synaptic activation during preparation¹⁴. After the preincubation period, the stimulation strength was determined for further testing by eliciting a population spike of 25% of its maximal amplitude. The stimulation intensity in the within-slice experiments was adjusted to obtain a field EPSP of 50% of its maximal slope in input S1 of populations 1 and 2, respectively. For S1 in population 1 and S2 in population 2, this intensity was the same strong synaptic activation as in experiments where the stimulus intensity was

determined by the mean of the population spike. Therefore, the stimulation intensity of S2 in population 1 varied between different experiments. The baseline was recorded for at least 60 min before LTP induction. Four 0.2-Hz biphasic, constant-current pulses (0.1 ms per polarity) were used for testing 1, 3, 5, 11, 15, 25, 30 min post-tetanus, and then once every 15 min. Late LTP was induced using three stimulus trains of 100 pulses (100 Hz, duration 0.2 ms per polarity) with 10 min intertrain intervals. This schedule of tetanization was used to produce very stable long-term maintenance of potentiation *in vitro* for longer than 8 h. In experiments in which a weaker induction of LTP was investigated, a short-term potentiation was induced using a single tetanus consisting of one train of 11 pulses (100 Hz, duration 0.1 ms per polarity, population spike threshold stimulus intensity for tetanization); and induction of early LTP alone using a single tetanus with 21 pulses (100 Hz, duration 0.2 ms per polarity, population spike threshold stimulus intensity for tetanization). The first weak tetanus resulted in a potentiation which lasted for ~1–2 h, whereas the second weak tetanization procedure caused a potentiation for ~3–5 h, unless preceded by late LTP to an independent pathway.

The responses were averaged and analysed on-line using a CED 1401 A/D converter 486/33 computer. The population spike amplitude (mV) was evaluated by taking the voltage difference between onset and peak. The slope function (mV per ms) of the field EPSP was calculated from 8 successive points of the steepest 400-µs segment.

Anisomycin, a reversible inhibitor of protein synthesis, was used at 25 µM, a concentration that blocks at least 85% of incorporation of ³H-leucine into hippocampal slices³. In addition to anisomycin, emetine, a second structurally different inhibitor, was used at 10 µM²⁶ to verify results in Fig. 2c, d (data not shown). Results were similar after the application of anisomycin either when emetine was applied throughout LTP induction of S1 and S2 ($n = 3$, field EPSP after 8 h in S1: $102.3 \pm 7.3\%$; S2: $98.9 \pm 15.8\%$) or only during induction of LTP on S2 ($n = 3$, field EPSP after 8 h in S1: $144 \pm 23.2\%$; S2: $165.3 \pm 32.0\%$). This result supports a specific action of the protein synthesis inhibitors. Drugs were added to the bath medium 25 min before LTP induction and were washed out 1 h after tetanization in the corresponding experiments. All drugs were dissolved in modified ACSF immediately before application.

Statistical analysis (Figs 2 and 3b, c) compared the time course of potentials between control LTP in input S1 and the low-frequency control input S2 in b or between control LTP in b and the potentiation of the corresponding groups using the two-tailed Mann-Whitney *U*-test ($P < 0.05$). In Fig. 3a the time course of the potentiation of population 1 was compared to the potentials of population 2 (Mann-Whitney *U*-test; $P < 0.05$). In Fig. 4, the time course of potentiation between LTP in input S1 and the potentiation in S2 was compared using the two-tailed Mann-Whitney *U*-test ($P < 0.05$).

Received 12 July; accepted 4 December 1996.

- Bliss, T. V. P. & Collingridge, G. L. *Nature* **361**, 31–39 (1993).
- Krug, M., Lössner, B. & Ott, T. *Brain Res. Bull.* **13**, 39–42 (1984).
- Frey, U., Krug, M., Reymann, K. G. & Matthies, H. *Brain Res.* **452**, 57–65 (1988).
- Fazeli, M. S., Errington, M. L., Dolphin, A. C. & Bliss, T. V. P. *Brain Res.* **473**, 51–59 (1988).
- Otani, S., Marshall, C. J., Tate, W. P., Goddard, G. V. & Abraham, W. C. *Neuroscience* **28**, 519–526 (1989).
- Fazeli, M. S., Corbett, J., Dunn, M. J., Dolphin, A. C. & Bliss, T. V. P. *J. Neurosci.* **13**, 1346–1353 (1993).
- Nguyen, P. V., Abel, T. & Kandel, E. R. *Science* **265**, 1104–1107 (1994).
- Frey, U., Frey, S., Schollmeier, F. & Krug, M. *J. Physiol.* **490**, 703–711 (1996).
- Link, W. et al. *Proc. Natl Acad. Sci. USA* **92**, 5734–5738 (1995).
- Davis, L., Banker, G. A. & Steward, O. *Nature* **330**, 477–479 (1987).
- Kleiman, R., Banker, G. & Steward, O. *Neuron* **5**, 821–830 (1990).
- Kang, H. & Schuman, E. M. *Science* **273**, 1402–1406 (1996).
- Huang, Y. Y. & Kandel, E. R. *Learning & Memory* **1**, 74–82 (1994).
- Frey, U., Schollmeier, K., Reymann, K. G. & Seidenbecher, T. *Neuroscience* **67**, 799–807 (1995).
- Steward, O. & Falk, P. M. *J. Neurosci.* **6**, 412–423 (1986).
- Lovinger, D. M. & Routtenberg, A. *J. Physiol. (Lond.)* **400**, 321–333 (1988).
- Stäubli, U. & Chun, D. *J. Neurosci.* **16**, 853–860 (1996).
- Hebb, D. O. *The Organization of Behaviour* (Wiley, New York, 1949).
- Larson, J. & Lynch, G. *Science* **232**, 985–988 (1986).
- Diamond, D. M., Dunwiddie, T. V. & Rose, G. M. *J. Neurosci.* **8**, 4079–4088 (1988).
- Malenka, R. C. *Neuron* **6**, 53–60 (1991).
- Abraham, W. C. & Bear, M. F. *Trends Neurosci.* **19**, 126–130 (1996).
- Rawlins, J. N. P. *Behav. Brain Sci.* 479–528 (1985).
- Squire, L. R. & Davis, H. P. *Annu. Rev. Pharmacol. Toxicol.* **21**, 323–356 (1981).
- Brown, R. & Kulik, J. *Cognition* **5**, 73–99 (1977).
- Stanton, P. K. & Sarvey, J. M. *J. Neurosci.* **4**, 3080–3088 (1984).

Acknowledgements. We thank D. Koch for her skilful technical assistance and D. Manahan-Vaughan for comments on the manuscript. This work was supported by a grant from the German BMBF to U.F.

Correspondence and requests for materials should be addressed to U.F. (e-mail: frey@ifn-magdeburg.de).

Integrin–ligand binding properties govern cell migration speed through cell–substratum adhesiveness

Sean P. Palecek*, Joseph C. Loftus†, Mark H. Ginsberg†, Douglas A. Lauffenburger* & Alan F. Horwitz‡

* Department of Chemical Engineering and Center for Biomedical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

† Department of Vascular Biology, The Scripps Research Institute, La Jolla, California 92037, USA

‡ Department of Cell and Structural Biology, University of Illinois, Urbana, Illinois 61801, USA

Migration of cells in higher organisms is mediated by adhesion receptors, such as integrins, that link the cell to extracellular-matrix ligands, transmitting forces and signals necessary for locomotion^{1–4}. Whether cells will migrate or not on a given substratum, and also their speed, depends on several variables related to integrin–ligand interactions, including ligand levels^{5,6}, integrin levels^{7–9}, and integrin–ligand binding affinities^{10–12}. These and other¹³ factors affect the way molecular systems integrate to effect and regulate cell migration. Here we show that changes in cell migration speed resulting from three separate variables—substratum ligand level, cell integrin expression level, and integrin–ligand binding affinity—are all quantitatively predictable through the changes they cause in a single unifying parameter: short-term cell–substratum adhesion strength. This finding is consistent with predictions of a mathematical model for cell migration¹⁴. The ligand concentration promoting maximum migration speed decreases reciprocally as integrin expression increases. Increases in integrin–ligand affinity similarly result in maximal migration at reciprocally lower ligand concentrations. The maximum speed attainable, however, remains unchanged as ligand concentration, integrin expression, or integrin–ligand affinity vary, suggesting that integrin coupling with intracellular motors remains unaltered.

Maximal cell migration speed is predicted to occur at an intermediate ratio of cell–substratum adhesiveness to intracellular contractile force, at which the cell can form new attachments at the cell front but break attachments at the rear^{6,14}. We therefore measured cell migration speed and short-term cell–substratum adhesiveness for different substratum ligand levels, integrin expression levels, and integrin–ligand binding affinities. This allowed us to test quantitatively the prediction that effects of changes in these molecular properties on cell migration speed can be predicted on the basis of their influence on short-term adhesiveness. Our short-term adhesion assay quantifies dynamic interactions between cell receptors and substratum ligands in a manner representative of events at the front and rear of migrating cells^{15,16}.

To vary receptor expression we transfected α_5 -deficient CHO B2 cells⁷ with a human α_5 cDNA, and sorted expressors by flow cytometry into populations with different relative expression levels of $\alpha_5\beta_1$ fibronectin receptors. To vary integrin affinity state we used CHO cells transfected with the $\alpha_{IIb}\beta_3$ ($K_a < 1.4 \times 10^4 \text{ M}^{-1}$) fibrinogen receptor and $\alpha_{IIb}\beta_3(\beta_{1-2})$ ($K_a = 4.85 \pm 0.84 \times 10^6 \text{ M}^{-1}$), a higher-affinity extracellular-domain mutation with 6 amino acids of the β_3 ligand-binding domain replaced with sequences derived from β_1 (ref. 17). Both $\alpha_{IIb}\beta_3$ and $\alpha_{IIb}\beta_3(\beta_{1-2})$ can be activated to higher affinity states by incubation with anti-LIBS2 antibody ($\alpha_{IIb}\beta_3$, $K_a = 1.66 \pm 0.33 \times 10^7 \text{ M}^{-1}$; $\alpha_{IIb}\beta_3(\beta_{1-2})$, $K_a = 4.55 \pm 0.77 \times 10^7 \text{ M}^{-1}$) (ref. 18). Mean CHO

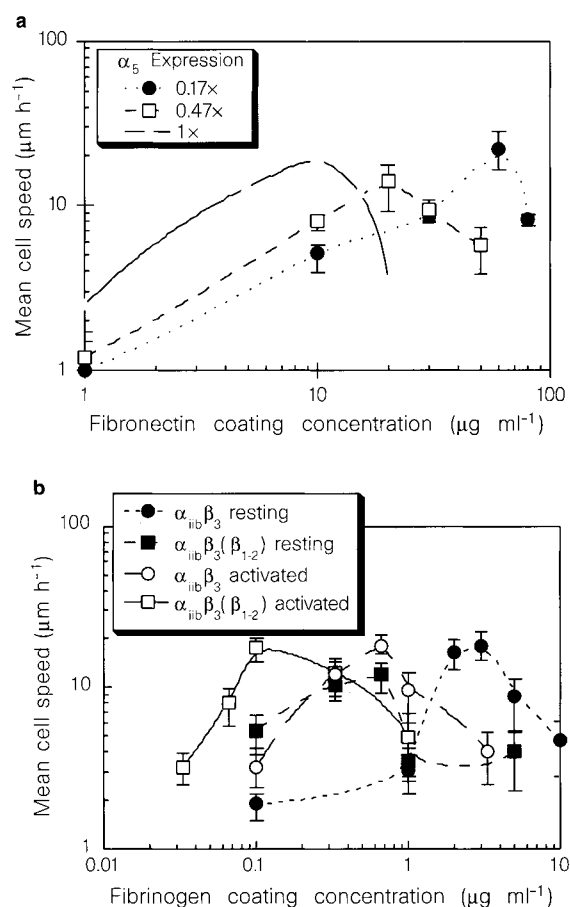


Figure 1 The speed of cell migration depends on substrate extracellular-matrix concentration, integrin expression level, and integrin–ligand affinity. Mean cell migration speed is determined using image analysis to track centroids of CHO cells with different integrin expression (**a**) and integrin–ligand affinity (**b**). Receptor expression was varied by transfecting α_5 -deficient CHO B2 cells with human α_5 and sorting populations with different expression. Integrin–ligand affinity was varied by transfecting CHO cells with $\alpha_{IIb}\beta_3$ and $\alpha_{IIb}\beta_3(\beta_{1-2})$, a higher-affinity mutant. These integrins were also activated by incubation with monoclonal antibody 62 (anti-LIBS2 antibody). Error bars represent 95% confidence intervals on the mean cell speed. CHO-cell populations with different integrin expression (**a**) or integrin–ligand affinity (**b**) all have a biphasic migration-speed dependence on fibronectin concentration. As α_5 expression (**a**) or $\alpha_{IIb}\beta_3$ –fibrinogen affinity increases (**b**), the extracellular-matrix concentration promoting maximal migration decreases; at low concentrations, migration speed increases as expression or affinity increases, whereas at high concentrations, migration speed decreases as expression or affinity increases. Maximum attainable migration speed is not a function of integrin expression or affinity.

cell migration speed, measured by time-lapse video-microscopy, exhibits a biphasic dependence on extracellular-matrix ligand concentration regardless of integrin expression level (the $\alpha_5\beta_1$ receptor on fibronectin; Fig. 1a) or integrin–ligand binding affinity (the $\alpha_{IIb}\beta_3$ receptor on fibrinogen; Fig. 1b). The maximum attainable cell speed remains the same in all cases; only the position of the biphasic curve is shifted. At low ligand concentrations, cell speed increases as integrin expression or binding affinity increases, whereas at high ligand concentrations, cell speed increases as integrin expression or binding affinity decreases. At intermediate ligand concentrations, intermediate integrin expression levels and intermediate binding-affinity states result in greatest cell speed. At maximum migration speeds (10–20 $\mu\text{m h}^{-1}$), CHO cells are moderately spread and move by extending multiple lamellae in different directions. Some

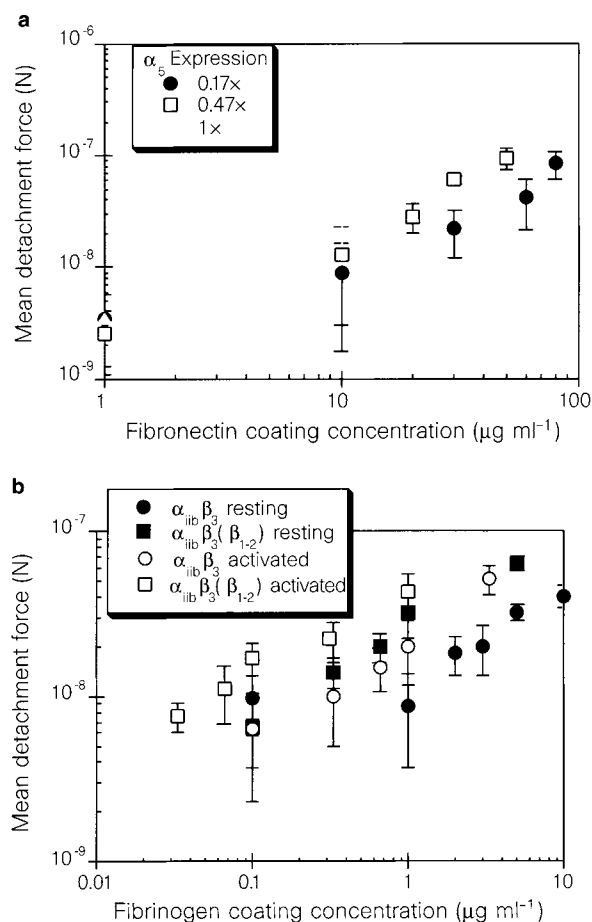


Figure 2 Cell-substratum adhesiveness increases as extracellular-matrix concentration, integrin expression, and integrin-ligand affinity increase. Short-term mean cell detachment force, measured by shear-flow detachment, was measured for CHO cells with different levels of integrin expression and integrin-ligand affinity. Receptor expression was varied by transfecting α_5 -deficient CHO B2 cells with human α_5 and sorting populations with different expression levels. Integrin-ligand affinity was varied by transfecting CHO cells with $\alpha_{5\text{ib}}\beta_3$ and $\alpha_{5\text{ib}}\beta_3(\beta_{1-2})$, a higher-affinity mutant. These integrins were also activated by incubation with monoclonal antibody 62 (anti-Lib2 antibody). Error bars represent 95% confidence intervals on the mean detachment force. In each cell population the mean detachment force increases with the extracellular-matrix concentration. At every concentration, the mean detachment force increases as integrin expression (a) or integrin-ligand affinity (b) increases.

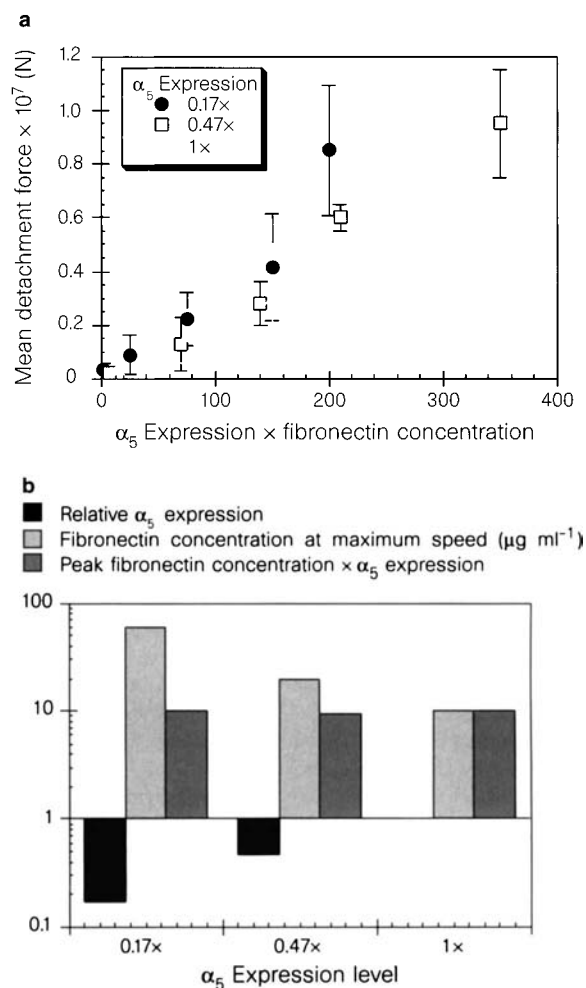


Figure 3 The number of cell-substratum bonds seems to determine cell-substratum adhesiveness and migration speed. At constant receptor-ligand affinity, the number of cell-substratum bonds is approximately linearly proportional to the product of extracellular-matrix concentration and receptor expression. Mean cell detachment force increases linearly with this product at each expression level (a). Thus short-term cell-substratum adhesiveness appears to be a linear function of the number of cell-substratum bonds. The product of α_5 expression and fibronectin concentration remains constant at the maximum migration speed (b), suggesting the existence of an optimal receptor occupancy at the maximum migration speed.

lamellae are stable and move the cell in the direction of their extensions, but other lamellae retract. At ligand concentrations lower than that giving rise to the maximum speed, the cells are more rounded but still extend lamellae, which are smaller and shorter lived than lamellae in migrating cells; these unstable lamellae cannot move the cell body. At ligand concentrations higher than that giving rise to the maximum speed, the cells are very spread and extend lamellae in a manner similar to migrating cells; the cell body does not move well, presumably because it cannot release adhesions to the substrate.

Short-term mean cell detachment force, measured by a shear-flow detachment assay¹⁹ after a 20-min incubation, increases with ligand concentration at each integrin expression level or affinity state (Fig. 2). At any ligand concentration, mean detachment force

increases as integrin expression or affinity increases. Increases in ligand concentration, integrin expression, or integrin affinity presumably allow the cells to form more bonds with the substrate, resulting in increased adhesiveness.

Integrin-ligand bond formation is assumed to be a second-order dynamic equilibrium process, such that $R + L \leftrightarrow C$, where L is the ligand, R the receptor, and C the receptor-ligand complex. Neglecting receptor and ligand depletion due to binding, $[C] = K'_a[R_0][L_0]$ so that the concentration of cell-substratum bonds, [C], is related to the product of the integrin concentration, $[R_0]$, and the extracellular-matrix ligand concentration, $[L_0]$, by an apparent association constant, K'_a . K'_a takes into account geometrical parameters of the immobilized ligand and the receptor diffusion in the plane of the cell membrane. At each α_5 expression level, mean detachment force

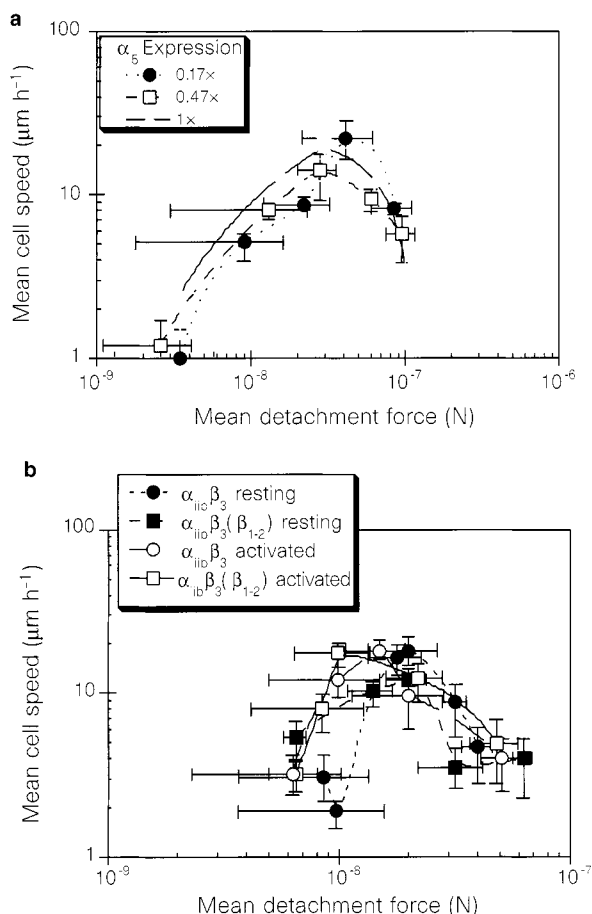


Figure 4 Cell migration speed (Fig. 1) correlates with cell-substratum adhesiveness (Fig. 2) as extracellular-matrix concentration, integrin expression, and integrin-ligand affinity change. Cell speed was measured by image analysis of cell centroids as a function of time, and detachment force was measured by shear-flow detachment. Error bars represent 95% confidence intervals on the mean cell speed and detachment force. CHO B2 cells transfected with human α_5 and sorted into populations with different relative α_5 expression demonstrate that, as fibronectin substratum concentration or α_5 expression change, the relationship between migration speed and mean detachment force is a constant, biphasic function (**a**). As $\alpha_{5\text{IIB}}\beta_3$ -fibrinogen affinity increases owing to the mutation from β_3 to $\beta_3(\beta_{1-2})$ and to activation induced by monoclonal antibody 62, the cell migration-speed dependence on short-term adhesiveness remains a constant, biphasic relationship (**b**). Maximum speed occurs at $2-4 \times 10^{-8}$ N for both $\alpha_5\beta_3$ - and $\alpha_{5\text{IIB}}\beta_3$ -mediated migration, suggesting that these receptors can interact with the cytoskeleton and transmit intracellular motile forces similarly as extracellular-matrix concentration, integrin expression, and integrin-ligand affinity vary.

has the same linear relationship to the product of α_5 expression and fibronectin concentration (Fig. 3a), suggesting that the short-term mean detachment force is proportional to the number of cell-substratum bonds. (At high ligand concentrations there may be significant receptor depletion, yielding an asymptotic plateau in this plot.) The fibronectin concentration allowing maximum migration speed is inversely proportional to the α_5 expression level, so that the product of the fibronectin concentration at maximum cell speed and the α_5 expression level is constant over the range of α_5 expression levels in this study (Fig. 3b). Because this product represents integrin-ligand bond number, it can be inferred that maximum migration occurs at an optimal, intermediate level of receptor occupancy.

To examine the relationship between migration speed and cell-

substratum adhesiveness, we replotted cell migration speed as a function of detachment force (Fig. 4). Cell migration speed is a constant, biphasic function of cell-substratum adhesiveness as integrin expression or integrin-ligand binding affinity changes. The presence of an optimum adhesiveness for cell migration indicates that modulation of ligand concentration, integrin expression, and integrin affinity can be used to alter cell migration speed through changes in the short-term adhesive interaction between the cell and substratum. Relatively small changes in integrin expression or affinity can lead to substantial changes in migration speed. These variables seem to contribute quantitatively to short-term cell adhesion by altering the number of cell-substratum bonds. Both the $\alpha_5\beta_1$ -fibronectin and the $\alpha_{5\text{IIB}}\beta_3$ -fibrinogen bonds promote maximum cell speed at a cell-substratum detachment force of $2-4 \times 10^{-8}$ N, suggesting that integrins with different ligand-recognition specificities can transmit intracellular motile forces similarly. Locomoting fibroblasts exert traction forces of about 20×10^{-8} N (ref. 20). These detachment forces and traction forces are consistent with the mathematical model prediction that maximum cell speed occurs at an adhesiveness to contractile force ratio of <1 (ref. 14). High cell-substratum adhesiveness probably hinders cell migration by obstructing the release of adhesions at the rear of the cell²¹⁻²³. Videotaped observations of cells at high adhesiveness reveal lamellipod extension and retraction, but little movement of the cell body. At low cell-substratum adhesiveness, however, inhibition of rear release is an unlikely explanation for reduced migration rates. Except at the lowest levels of adhesiveness, the cells are still able to extend lamellae, but these lamellae are smaller and less frequently extended than those in more adherent cells. Decreased rate of lamellipodal extension or more likely, decreased probability of formation of a stable attachment to the surface by the lamellae, probably account for the reduced migration at low adhesiveness²⁴.

Additional variables such as lamellipodal extension, intracellular force generation, integrin clustering and avidity effects, and integrin signalling are important to the processes of cell adhesion and cell migration^{4,12,13}. For example, overexpression of FAK in CHO cells results in increased cell migration but does not alter cell spreading or adhesion, at least as measured in a particular assay²⁵. Nonetheless, our results suggest that short-term cell-substratum adhesiveness is rate limiting in determining migration speed as the linkage between integrin and the extracellular matrix is altered, and can serve as a unifying parameter in understanding the influence of various molecular properties on migration speed.

This interrelation between migration and adhesiveness has important implications in clinical applications, such as various molecular therapeutic approaches to cancer²⁶. Where decreased cell migration is generally desirable, targeting receptor expression or receptor-ligand binding affinity are potentially effective strategies. In biomaterials and tissue-engineering applications, increased migration is often desired²⁷. Altering the extracellular-matrix ligand concentration seems to be the easiest means of affecting cell-substratum adhesiveness, but other approaches to change receptor number or receptor-ligand binding affinity, such as soluble binding competitors or drugs affecting receptor-ligand or receptor-cytoskeleton coupling, could allow additional control of cell adhesiveness and migration in a quantitatively predictable manner. □

Methods

Cell preparation. CHO B2 cells⁷ were transfected with a human α_5 cDNA²⁸ using lipofectamine (GIBCO-BRL), as described in the manufacturer's protocols. Generation of CHO cells expressing $\alpha_{5\text{IIB}}\beta_3$ and $\alpha_{5\text{IIB}}\beta_3(\beta_{1-2})$ has been described¹⁷. The cells were grown in DMEM containing 10% FBS, 2 mM glutamine, and 1% non-essential amino acids. The transfected cells were selected in DME containing $100 \mu\text{g ml}^{-1}$ G418 and maintained in DME containing $50 \mu\text{g ml}^{-1}$ G418. Cells expressing human α_5 were selected

from non-expressors by flow cytometry. Quantitative cell-surface expression of integrins was determined by flow cytometry²⁹ using human α_5 antibody 6F4 at 1:4 dilution of hybridoma supernatant to assay α_5 expression and non-inhibitory $\alpha_{IIb}\beta_3$ antibody D57 at 1:200 dilution of mouse ascites to assay $\alpha_{IIb}\beta_3$ expression. Cells expressing α_5 were sorted into three populations with different relative expression levels, and cells expressing $\alpha_{IIb}\beta_3(\beta_{1-2})$ were sorted into similar surface-expression profiles to $\alpha_{IIb}\beta_3$ transfected cells. Both $\alpha_{IIb}\beta_3$ and $\alpha_{IIb}\beta_3(\beta_{1-2})$ were activated to the high-affinity state by incubation with 100 μgml^{-1} monoclonal antibody 62 (anti-LIBS2 antibody).

Adhesion assay. CHO cells were incubated on silanated fibrinogen- or fibronectin (Sigma)-coated glass slides for 20 min in serum-free OptiMEM 1 (GIBCO-BRL). The slide was placed in a shear-stress flow chamber¹⁹, which produces a linear gradient in shear with position along the centre line: $\tau_w = (6\mu Q/h^2 w_1)/(1 - (z/L))$ where τ_w is the surface shear stress, ν the fluid viscosity, Q the flow rate, h the channel height, w_1 the channel width at the origin of the flow field, z the distance from the origin along the centre line, and L the length of the flow field. PBS with Ca^{2+} and Mg^{2+} , heated to 37 °C, flowed through the chamber for 5 min to detach the cells. Cells in 20 fields along the slide were counted before and after flow detachment. Shear stress was calculated for each field and converted to shear force (F_s) by approximating cell morphology as a hemispherical cap, such that $F_s = 2.15\pi(r_p^2 + h^2)\tau_w$, where r_p is the hemispherical cap radius and h the height. The fraction of cells detached as a function of shear force was fit to the integrals of logarithmic normal probability density-function distributions to determine the mean shear force for detachment of 50% of the cells. Five detachment assays were performed for each cell type at each extracellular-matrix concentration, and mean detachment forces were averaged.

Migration assay. CHO cells were incubated on silanated, fibrinogen- or fibronectin-coated coverslips for 3 h in serum-free OptiMEM 1. Real-time digital image processing was used to acquire images and calculate cell centroid position as a function of time. The image-processing software (Engineering Technology Center, Mystic, CT) identifies cell boundaries from phase-contrast images and measures cell centroid position. We scanned 5–10 cells per field in 10 different fields every 15 min for 12 h. The mean-squared displacement of the cell centroid as a function of time was calculated for each cell using non-overlapping time intervals. The mean-squared displacements were averaged and fit to a persistent random walk model³⁰ to calculate cell speed, S , and persistence time, P : $\langle d^2(t) \rangle = 2S^2 P [t - P(1 - e^{-t/P})]$.

Received 13 November; accepted 19 December 1996.

- Hynes, R. O. *Cell* **69**, 11–25 (1992).
- Cheresh, D. A. *Adv. Mol. Cell Biol.* **6**, 225–252 (1993).
- Schwartz, M. A., Schaller, M. D. & Ginsberg, M. H. *Annu. Rev. Cell Dev. Biol.* **11**, 549–599 (1995).
- Lauffenburger, D. A. & Horwitz, A. F. *Cell* **84**, 359–369 (1996).
- Goodman, S. L., Risse, G. & van der Mark, K. J. *Cell Biol.* **109**, 799–809 (1989).
- DiMilla, P. A., Stone, J. A., Quinn, J. A., Albelda, S. A. & Lauffenburger, D. A. *J. Cell Biol.* **122**, 729–737 (1993).
- Bauer, J. S., Schreiner, C. L., Giancotti, F. G., Ruoslahti, E. & Juliano, R. L. *J. Cell Biol.* **116**, 477–487 (1992).
- Giancotti, F. G. & Ruoslahti, E. *Cell* **60**, 849–859 (1990).
- Keely, P. J., Fong, A. M., Zutter, M. M. & Santoro, S. A. *J. Cell Sci.* **108**, 595–607 (1995).
- Duband, J.-L., Dufour, S., Yamada, S. S., Yamada, K. M. & Thiery, J. P. *J. Cell Sci.* **98**, 517–532 (1991).
- Kuipers, T. W. et al. *J. Exp. Med.* **178**, 279–284 (1993).
- Huttenlocher, A., Ginsberg, M. H. & Horwitz, A. F. *J. Cell Biol.* **134**, 1551–1562 (1996).
- Huttenlocher, A., Sandborg, R. R. & Horwitz, A. F. *Curr. Opin. Cell Biol.* **7**, 697–706.
- DiMilla, P. A., Barbee, K. & Lauffenburger, D. A. *Biophys. J.* **60**, 15–37 (1991).
- Regen, C. M. & Horwitz, A. F. *J. Cell Biol.* **119**, 1347–1359 (1992).
- Palecek, S. P., Schmidt, C. E., Lauffenburger, D. A. & Horwitz, A. F. *J. Cell Sci.* **109**, 941–952 (1996).
- Bajt, M. L., Loftus, J. C., Gawaz, M. P. & Ginsberg, M. H. *J. Biol. Chem.* **267**, 22211–22216 (1992).
- Frelinger, A. L., Du, X., Plow, E. F. & Ginsberg, M. H. *J. Biol. Chem.* **266**, 17106–17111 (1991).
- Powers, M. J., Rodriguez, R. E. & Griffith, L. G. *Biotechnol. Bioeng.* (in the press).
- Oliver, T. N., Lee, J. & Jacobson, K. *Semin. Cell Biol.* **5**, 139–147 (1994).
- Abercrombie, M., Heaysman, J. E. M. & Pegrun, S. M. *Exp. Cell Res.* **59**, 393–398 (1970).
- Marks, P. W., Hendey, B. & Maxfield, F. R. *J. Cell Biol.* **112**, 149–158 (1991).
- Jay, P. Y., Pham, P. A., Wong, S. A. & Elson, E. L. *J. Cell Sci.* **108**, 387–393 (1995).
- Wessels, D., Vawter-Hugart, H., Murray, J. & Soll, D. R. *Cell Motil. Cytoskeleton* **27**, 1–12 (1994).
- Cary, L. A., Chang, J. F. & Guan, J. J. *Cell Sci.* **109**, 1787–1794 (1996).
- Varner, J. A. & Cheresh, D. A. *Curr. Opin. Cell Biol.* **8**, 724–730 (1996).
- Langer, R. & Vacanti, J. *Science* **260**, 920–926 (1993).
- Argaves, W. S. et al. *J. Cell Biol.* **105**, 1183–1190 (1987).
- Loftus, J. C. et al. *Science* **249**, 915–918 (1990).
- Dunn, G. A. *Agents Actions* (suppl.) **22**, 14–33 (1983).

Acknowledgements. We thank R. Isberg for 6F4 antibody, L. Reichardt for α_5 cDNA, and R. Juliano for CHO B2 cells. This work was supported by grants from the NIH to D.A.L., A.E.H., J.C.L. and M.H.G., and a Whitaker Foundation graduate fellowship in Biomedical Engineering to S.P.P.

Correspondence and requests for materials should be addressed to D.A.L. (e-mail: lauffen@mit.edu).

MAP3K-related kinase involved in NF- κ B induction by TNF, CD95 and IL-1

Nikolai L. Malinin, Mark P. Boldin, Andrei V. Kovalenko & David Wallach

Department of Membrane Research and Biophysics, The Weizmann Institute of Science, 76100 Rehovot, Israel

Several members of the tumour-necrosis/nerve-growth factor (TNF/NGF) receptor family activate the transcription factor NF- κ B through a common adaptor protein, Traf2 (refs 1–5), whereas the interleukin 1 type-I receptor activates NF- κ B independently of Traf2 (ref. 4). We have now cloned a new protein kinase, NIK, which binds to Traf2 and stimulates NF- κ B activity. This kinase shares sequence similarity with several MAPKK kinases. Expression in cells of kinase-deficient NIK mutants fails to stimulate NF- κ B and blocks its induction by TNF, by either of the two TNF receptors or by the receptor CD95 (Fas/Apo-1), and by TRADD, RIP and MORT1/FADD, which are adaptor proteins that bind to these receptors. It also blocked NF- κ B induction by interleukin-1. Our findings indicate that NIK participates in an NF- κ B-inducing signalling cascade common to receptors of the TNF/NGF family and to the interleukin-1 type-I receptor.

NF- κ B, a ubiquitously expressed transcription factor comprising a homo- or heterodimer of DNA-binding proteins related to the proto-oncogene c-Rel, controls the expression of many immune- and inflammatory-response genes. In most cells NF- κ B exists in a

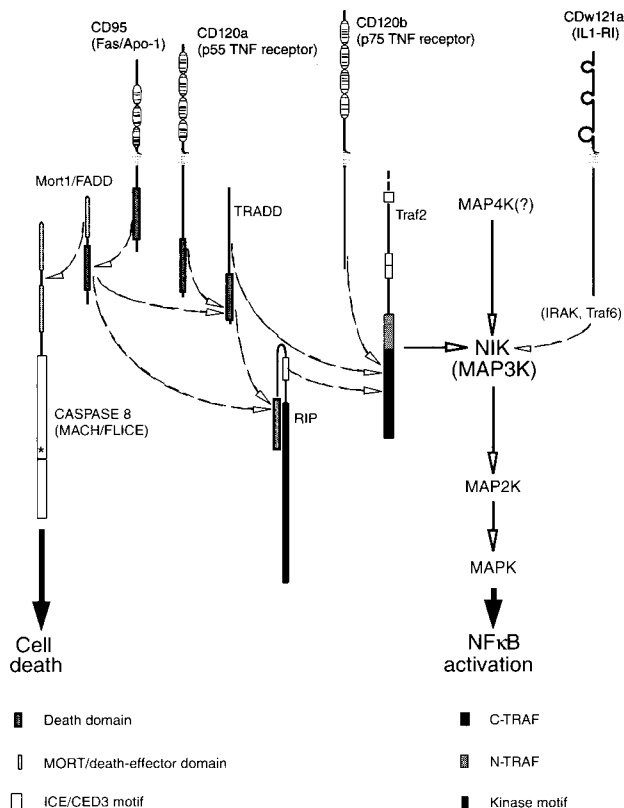


Figure 1 The known protein-protein interactions through which the TNF receptors (p55, or CD120a, and p75, or CD120b), CD95 (Fas/Apo-1) and CDw121a (the IL-1 type-I receptor) might affect the activity of NIK.

latent state in the cytoplasm bound to inhibitory proteins (collectively called I κ B) that mask its nuclear localization signal. The latent form can be activated by inducing agents, including several cytokines that signal for phosphorylation and subsequent degradation of I κ B (reviewed in ref. 6). Several adaptor proteins are involved in the events associated with this cytokine-induced effect, including the TRAF proteins (for TNF-receptor-associated factors). These proteins share sequence homology at their C-terminal receptor-binding regions (TRAF domain), although their binding properties and activities differ. One of them, Traf2, binds to the p55 (CD120a) and p75 (CD120b) TNF receptors, as well as to several other receptors of the TNF/NGF receptor family, either directly or through other adaptor proteins (Fig. 1). Traf2 is crucial for the activation of NF- κ B by these receptors⁵; Traf3 inhibits activation of NF- κ B by some receptors of the TNF/NGF family², whereas Traf6 is required for the induction of NF- κ B by interleukin-1 (IL-1)⁷, a cytokine with activities similar to TNF, although it uses a structurally unrelated receptor. None of the TRAF molecules has enzymatic activity.

Here we describe a new protein with serine/threonine kinase activity that binds to Traf2 and takes part in the activation of NF- κ B by the TNF receptors, by CD95 (also known as Fas/Apo-1, a receptor belonging to the TNF/NGF receptor family), and by the type-I IL-1 receptor (CDw121a). This protein has marked sequence similarity to several kinases that participate in MAP kinase cascades and may itself act in a kinase cascade.

A partial complementary DNA of this new protein was cloned by two-hybrid screening of a human B-cell cDNA library^{8,9} using Traf2 as bait. Employing several steps of the polymerase chain reaction (PCR), we then cloned the full-length cDNA from cDNA libraries. This cDNA was 4,596 nucleotides long, with an open reading frame of 2,841 nucleotides, of which 972 were present in the initially cloned partial cDNA. In view of the kinase activity of the protein and its ability to stimulate NF- κ B (see below), we called it NIK, for NF- κ B-inducing kinase.

Two-hybrid testing of the binding properties of NIK (Fig. 2) revealed that the initially isolated partial clone of NIK (NIK624–947) binds specifically to the C-terminal region (TRAF domain) of Traf2. By contrast, binding of full-length NIK to Traf2 involved both the TRAF domain and the region upstream of it. NIK did not bind to Traf3. Moreover, a chimaeric molecule comprising the TRAF domain of Traf2 and the N-terminal part of Traf3 bound NIK624–947 but not full-length NIK, confirming that binding of full-length NIK to Traf2 involves both the N- and C-terminal regions of Traf2. NIK did not self-associate, neither did it bind to the intracellular domain of either of the two TNF receptors, nor to CD40 or CD95 (two receptors that belong to the TNF/NGF family). It did not bind to TRADD, MORT1/FADD or RIP, which share a motif (the 'death' domain) that differs from the TRAF domain. These three adaptor proteins are involved in the function of the TNF receptors and of CD95 and trigger induction of NF- κ B (see below), apparently by interacting with Traf2 (Fig. 1).

NIK also bound Traf2 in transfected mammalian cells, where its expression was markedly increased upon deletion of its 3' non-coding region (Fig. 3).

Northern blot analysis revealed a single transcript of NIK, expressed in various tissues at different and rather low levels, whose size (~5,000 nucleotides) was similar to that of cloned NIK cDNA (Fig. 4a).

Examination of the predicted amino-acid sequence of NIK (Fig. 4b) revealed that this protein contains a serine/threonine protein-kinase motif, resembling several MAP kinase kinases (MAPKKK³; Fig. 4c). *In vitro* testing of NIK kinase activity indicated that the protein can be autophosphorylated, but not when the active-site lysine and an adjacent lysine are replaced with alanine (NIK KK429–430AA in Fig. 5).

Overexpression of NIK in 293 EBNA cells induced NF- κ B, and to

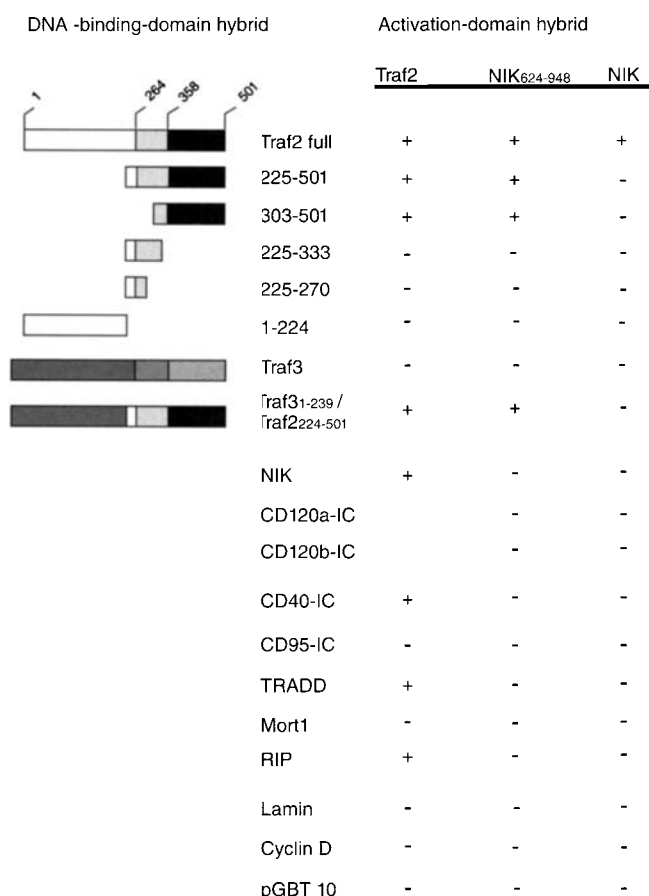


Figure 2 Interaction of NIK with Traf2 in transfected yeast. The binding of NIK and its partial clone NIK624–947 to Traf2, Traf2 deletion mutants, Traf3, a chimaera of Traf3 (residues 1–239) and Traf2 (residues 224–501), and to specificity controls encoded in transfected SFY526 yeast by the Gal4 DNA-binding-domain and activation-domain constructs (pGBT10 and pGAD-GH) was assessed by two-hybrid β -galactosidase filter assay¹⁵. Plus and minus signs indicate that strong colour develops within 3 h and no colour develops within 24 h, respectively. IC, intracellular domain.

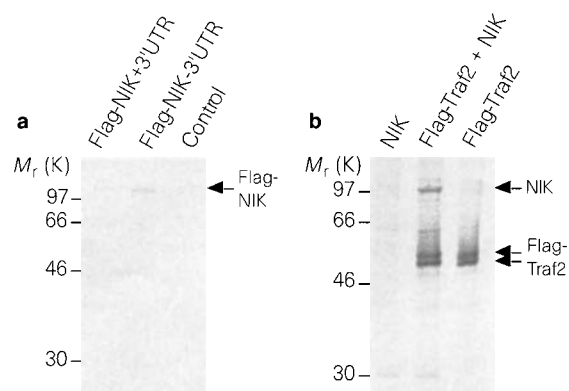


Figure 3 Interaction of NIK with Traf2 within transfected 293-EBNA cells. **a**, Effect of the NIK 3' untranslated region (UTR) on its expression. NIK was expressed and metabolically labelled with ³⁵S in 293-EBNA cells, with or without its 3' UTR, as a fusion protein with the Flag epitope. Immunoprecipitation of the fusion protein using anti-Flag antibody is shown. **b**, Interaction of NIK with Traf2, showing the immunoprecipitation of Traf2 fused at its N terminus to the Flag epitope (Flag-Traf2), which was expressed alone or with NIK and metabolically labelled.

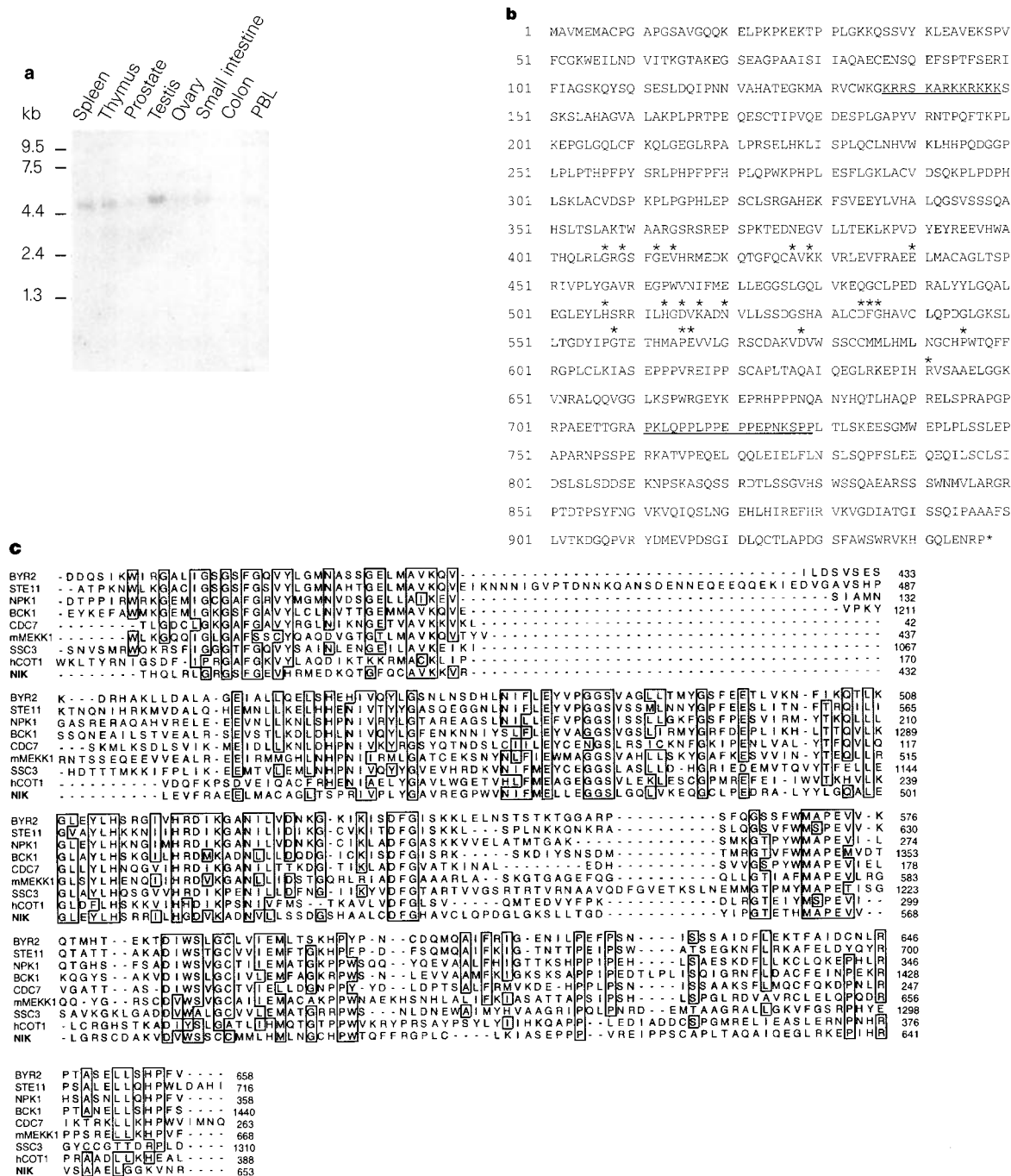


Figure 4 NIK transcript and its encoded protein. **a**, Northern blot of poly(A)⁺ RNA (2 µg per lane) from human tissues (Clontech) probed with a cDNA corresponding to nucleotides 1,259–2,150 in NIK. PBL, peripheral blood leukocytes. **b**, Deduced amino-acid sequence of NIK. Conserved residues in the kinase domain¹⁹ are indicated by asterisks. A proline-rich region downstream of the kinase domain and an upstream lysine-rich region (which may act as a nuclear localization signal) are underlined. **c**, Collinear amino-acid sequence alignment of

the kinase domains in NIK, the mouse MEKK1 (ref. 20) and the proteins of cluster 2032.9 in the CRSeqAnnot database²¹ which includes Ste11 (ref. 22), Bck1 (ref. 23) and Ste8 (also called Byr2; refs 24, 25), yeast proteins that serve as MAPKKK, the yeast proteins Cdc7 (ref. 26) and Ssc3 (SwissProt P25390), the mammalian oncogene *cot* (also called the Ewing's sarcoma oncogene or Tpl-2; refs 27, 28) which acts as MAPKKK²⁹, and the tobacco Ste11 homologue NPK1 (ref. 30).

a greater extent than did overexpression of Traf2. However, NF-κB was not activated in cells expressing NIK624–947 or NIK KK429–430AA (Fig. 6a, b), indicating that induction by the native molecule depends on its kinase function. NIK differs in this respect from RIP, which also contains a protein kinase domain yet can still induce NF-κB when its kinase activity is abolished by mutation (Fig. 6b).

The activation of NF-κB upon overexpression of NIK was indistinguishable from that produced by treatment of cells with TNF. As with TNF or Traf2 (ref. 2) overexpression, the principal components of NIK-activated NF-κB were p50 and p65 (Fig. 7a). NIK over-

expression caused degradation of IκBα, and blocking this degradation by N-acetyl-Leu-Leu-norleucinal (ALLN) resulted, as with TNF, in the accumulation of IκB molecules that migrated more slowly in SDS–PAGE and might correspond to phosphorylated IκBα (Fig. 7b, c).

As shown in Fig. 6, NF-κB can be activated in 293-EBNA cells by TNF, as well as by overexpression of CD120a or CD120b, or of CD120a with its intracellular domain replaced by that of CD95 (CD120a-CD95). It can also be activated by overexpression of Traf2, TRADD, RIP or MORT1/FADD, but not by a MORT1/FADD

deletion mutant lacking the region upstream of the death domain. Expression of NIK KK429–430AA or of NIK624–947 in 293-EBNA cells blocked induction of NF- κ B by all of these agents, indicating that NIK activity is involved in the induction (Fig. 6). This inhibition correlated with less I κ B reduction (Fig. 7c).

NF- κ B is activated by IL-1 as well. This effect, unlike that of the TNF receptors and CD95, is little affected by expression of a Traf2 dominant-negative mutant⁴ (Fig. 6b, right), but is inhibited by expression of NIK mutants (Fig. 6b, left). However, the NF- κ B activity observed upon overexpression of the p65 Rel homologue in 293-EBNA cells was unaffected by coexpression of kinase-deficient NIK mutants, indicating that NIK does not affect the function of Rel proteins directly, but participates in their receptor-induced activation.

The cytotoxic activity of TNE, which is apparently mediated by a Mort1/FADD-associated ICE/Ced3 protease (caspase 8/MACH/

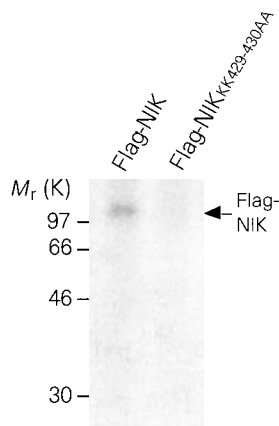


Figure 5 NIK kinase activity. Autoradiography of SDS-PAGE to assess the *in vitro* autophosphorylation of Flag-NIK and Flag-NIK KK429–430AA, immunoprecipitated from lysates of 293-EBNA cells that transiently express these proteins.

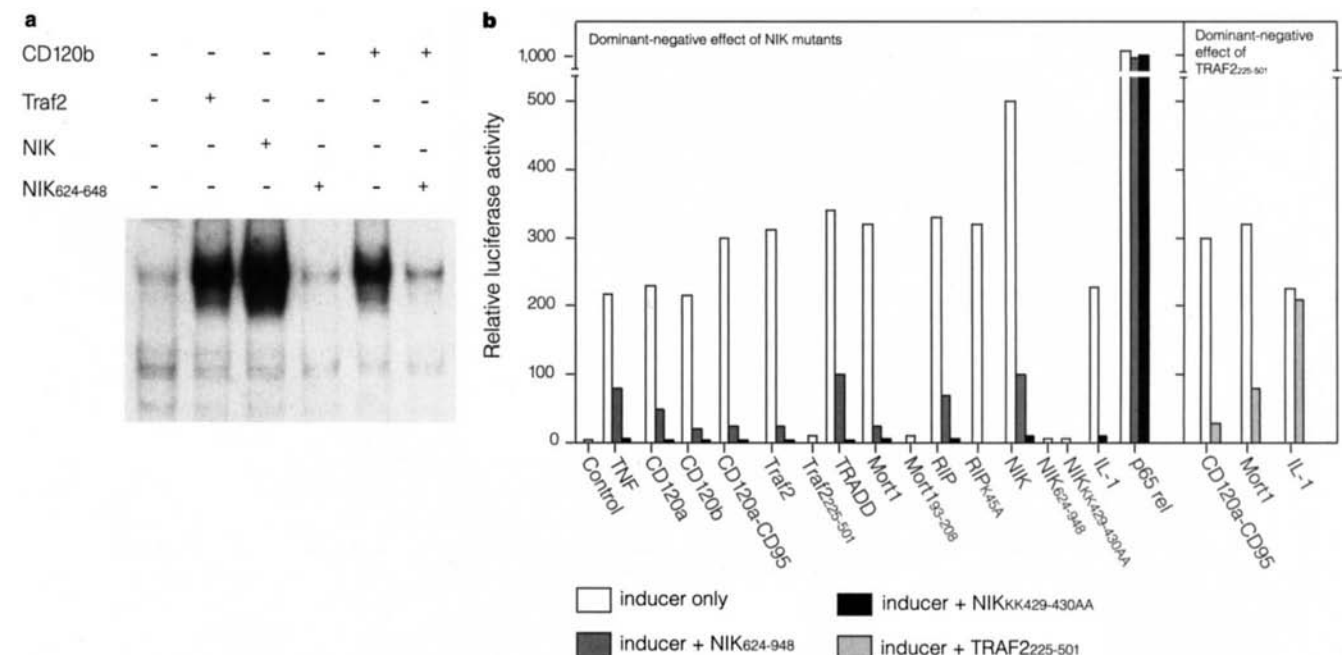


Figure 6 NIK mediates NF- κ B induction by the TNF receptors, CD95 and IL-1. **a**, Electromobility shift assay of NF- κ B binding activity demonstrating the activation of NF- κ B by overexpression of Traf2, NIK or CD120b in 293-EBNA cells, and inhibition of the CD120b-induced effect by coexpression of the kinase-deficient partial clone NIK624–947. **b**, NF- κ B-dependent reporter gene activity (HIV LTR-luciferase) in 293-EBNA cells treated with TNF (300 U ml⁻¹) or IL-1 α (300 U m⁻¹) for

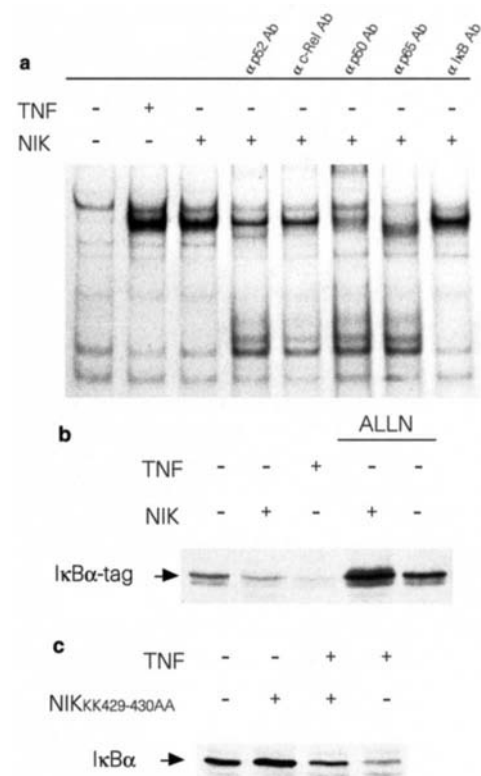


Figure 7 Rel and I κ B molecules affected by NIK. **a**, Effects of antibodies against different Rel homologues on the NIK-induced NF- κ B complex. Antisera against Rel proteins or, as a control, against I κ B α were added in aliquots of 1 μ l per electromobility-shift assay reaction. **b** and **c**, Western blot analysis of I κ B α expression showing degradation of I κ B α or, in cells treated with *N*-acetyl-Leu-Leu-norleucinal (ALLN; 100 mM for 3 h), accumulation of phosphorylated I κ B α , in response to NIK expression (**b**), and inhibition of the degradation of I κ B α , induced by TNF (3,000 U m⁻¹ for 30 min) in cells expressing NIK KK429–430AA (**c**).

6 h, or in 293-EBNA cells overexpressing the TNF receptors, CD120a–CD95 or their adaptor proteins, or p65 NF- κ B subunit; effects of coexpression of NIK KK429–430AA or NIK624–947 (left panel), or Traf2 225–501 (right panel). Data are from one of five experiments with similar qualitative results. They are presented in terms of luciferase activity relative to that in cells transfected with empty vector.

FLICE; Fig. 1), is subject to negative regulation by some NF- κ B-inducible genes (see ref. 10 for example). The antagonizing consequences of NF- κ B-mediated gene induction and caspase 8/MACH/FLICE activation may explain why TNF itself, as well as IL-1 (both activate NF- κ B), can induce cellular resistance to TNF cytotoxicity^{11,12}. In line with this, we found that expression of NIK dominant-negative mutants in 293-EBNA cells significantly increased their susceptibility to killing by TNF, and that overexpression of wild-type NIK inhibited their killing by TNF or as a result of overexpression of CD120a (data not shown).

Our findings indicate that the kinase activity of NIK is part of a signalling cascade that is responsible for NF- κ B activation and is common to the TNF receptors CD95 and IL-1, although its exact role in this process is not clear. Binding of NIK to Traf2 may allow NIK to be affected by both TNF receptors and by CD95 (refs 2,4; Fig. 1). By analogy with the MAP kinase cascades, NIK may serve as substrate for a kinase (MAPKKKK) upon being recruited with Traf2 to the stimulated receptors, so that when it is phosphorylated it phosphorylates and hence activates other kinase(s). The IL-1-induced activation of NF- κ B, however, is independent of TRAF2. Perhaps activation of NIK by the IL-1 receptor is mediated by IRAK, a serine/threonine kinase that is recruited to the IL-1 receptor after stimulation¹³, and by Traf6, which binds IRAK⁷. The target of NIK, or of a cascade of kinases activated by it, might be I κ B, whose phosphorylation facility is unknown. Alternatively, NIK may phosphorylate TRAF proteins or regulatory proteins that bind to them, for example TANK/I-TRAF^{3,14}, creating docking sites for other proteins. Identification of the cellular substrates of NIK will shed more light on its mechanism of action and its functional relationship to known MAP kinase cascades. □

Methods

Metabolic labelling, immunoprecipitation and kinase assay. Proteins were expressed by transfecting pcDNA3 expression constructs with the indicated cDNAs into human embryonic kidney 293-EBNA cells (5×10^6 cells per sample). Extracts were prepared 24 h after transfection. Except where otherwise indicated, NIK cDNA was expressed without its 3' UTR. Metabolic labelling and immunoprecipitation were done as described¹⁵, using cells transfected with 10 μ g NIK expression vector. For analysis of NIK autophosphorylation, cells were lysed in 25 mM Tris-HCl, pH 5.7, 150 mM NaCl, 1% Triton X-100, 1 mM EDTA, 1 mM EGTA and 20 mM NaF. Flag-tagged proteins were immunoprecipitated with anti-Flag affinity gel and then incubated for 20 min at 30 °C in a buffer containing 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 6 mM MgCl₂, 1 mM MnCl₂, 1 mM DTT and 100 mM [γ -³²P]ATP (150 mCi μ mol⁻¹), followed by SDS-PAGE.

NF- κ B activity measurements. Electromobility-shift assay was done as described¹⁶, using a radiolabelled double-stranded oligonucleotide of sequence 5'-GATGCCATTGGGGATTCTCTTTGAC-3'. For the reporter gene assay, cells were cotransfected with the HIV LTR-luciferase reporter gene plasmid (1 μ g) and the NF- κ B-inducing-protein (1.5 μ g) and NIK-mutant (3 μ g) expression vectors. The amount of transfected DNA was kept constant by supplementation with 'empty' vector. Overexpression of kinase-deficient NIK mutants did not affect the activity of coexpressed luciferase linked to the β -actin promoter (not shown), ruling out the possibility that inhibition of NF- κ B induction by the mutants reflects a cytotoxic effect.

Western blot analysis of I κ B expression. Cells were transfected with expression vectors of NIK or NIK KK429-430AA or with empty vector (3 μ g) and, for the experiment shown in Fig. 7b, with 1.5 μ g of an expression vector for C-terminally tagged I κ B α (pcDNA3-I κ B ctag¹⁷) as well. Cell extracts were analysed by SDS-PAGE (12%) followed by western blot analysis, using the anti-Tag monoclonal antibody SV5-Pk¹⁸ (Fig. 7b), or rabbit anti-I κ B α antiserum (Fig. 7c), and the ECL kit (Amersham).

Received 14 November; accepted 23 December 1996.

1. Rothe, M., Wong, S. C., Henzel, W. J. & Goeddel, D. V. *Cell* **78**, 681-692 (1994).
2. Rothe, M., Sharma, V., Dixit, V. M. & Goeddel, D. V. *Science* **269**, 1424-1427 (1995).
3. Cheng, G. & Baltimore, D. *Genes Dev.* **10**, 963-973 (1996).
4. Hsu, H., Shu, H.-B., Pan, M.-G. & Goeddel, D. V. *Cell* **84**, 299-308 (1996).
5. Wallach, D. *Eur. Cytokine Net.* **7**, 713-724 (1996).

6. Miyamoto, S. & Verma, I. M. *Adv. Cancer Res.* **66**, 255-292 (1995).
7. Cao, Z., Xiong, J., Takeuchi, M., Kurama, T. & Goeddel, D. V. *Nature* **383**, 443-446 (1996).
8. Durfee, T. et al. *Genes Dev.* **7**, 555-569 (1993).
9. Boldin, M. P., Goncharov, T. M., Goltsev, Y. V. & Wallach, D. *Cell* **85**, 803-815 (1996).
9. Seger, R. & Krebs, E. G. *FASEB J.* **9**, 726-735 (1995).
10. Beg, A. A. & Baltimore, D. *Science* **274**, 782-784 (1996).
11. Hahn, T. et al. *Proc. Natl Acad. Sci. USA* **82**, 3814-3818 (1985).
12. Holtmann, H., Hahn, T. & Wallach, D. *Immunobiology* **177**, 7-22 (1988).
13. Cao, Z., Henzel, W. J. & Gao, X. *Science* **271**, 1128-1131 (1996).
14. Rothe, M. et al. *Proc. Natl Acad. Sci. USA* **93**, 8241-8246 (1996).
15. Boldin, M. P. et al. *J. Biol. Chem.* **270**, 7795-7798 (1995).
16. Schreiber, E., Matthias, P., Muller, M. M. & Schaffner, W. *Nucleic Acids Res.* **17**, 6419 (1989).
17. Rodriguez, M. S., Michalopoulos, I., Arenzana-Seisdedos, F. & Hay, R. T. *Mol. Cell. Biol.* **15**, 2413-2419 (1995).
18. Hanke, T., Szwilowski, P. & Randall, R. E. *J. Gen. Virol.* **73**, 653-660 (1992).
19. Hanks, S. K., Quinn, A. M. & Hunter, T. *Science* **241**, 42-52 (1988).
20. Lange-Carter, C. A., Fleiman, C. M., Gardner, A. M., Blumer, K. J. & Johnson, G. L. *Science* **260**, 315-319 (1993).
21. Worley, K. C., Wiese, B. A. & Smith, R. F. *Genome Res.* **5**, 173-184 (1995).
22. Rhodes, N., Connell, L. & Errede, B. *Genes Dev.* **4**, 1862-1874 (1990).
23. Lee, K. S. & Levin, D. E. *Mol. Cell. Biol.* **12**, 172-182 (1992).
24. Wang, Y., Xu, H. P., Riggs, M., Rodgers, L. & Wigler, M. *Mol. Cell. Biol.* **11**, 3554-3563 (1991).
25. Styrkarsdottir, U., Egel, R. & Nielsen, O. *Mol. Gen. Genet.* **235**, 122-130 (1992).
26. Fankhauser, C. & Simanis, V. *EMBO J.* **13**, 3011-3019 (1994).
27. Miyoshi, J., Higashi, T., Mukai, H., Ohuchi, T. & Kakunaga, T. *Mol. Cell. Biol.* **11**, 4088-4096 (1991).
28. Patriotic, C., Makris, A., Bear, S. E. & Tschlis, P. N. *Proc. Natl Acad. Sci. USA* **90**, 2251-2255 (1993).
29. Salmeron, A. et al. *EMBO J.* **15**, 817-826 (1996).
30. Banno, H. et al. *Mol. Cell. Biol.* **13**, 4745-4752 (1993).

Acknowledgements. We thank R. Seger for helpful comments; J.-L. Virelizier, Y. Ben-Neriah and I. Alkalay for their advice and for reagents used in the assessment of NF- κ B activation; S. J. Elledge for the B-cell cDNA library; and H. Englemann for CD40 cDNA. This work was supported in part by grants from InterLab Ltd, Ness Ziona, Israel, from Ares Trading SA, Switzerland, from the Israeli Ministry of Arts and Sciences and from the Forchheimer Center for Molecular Genetics.

Correspondence and requests for materials should be addressed to D.W. (e-mail: lwalach@weizmann.weizmann.ac.il). Nucleotide and amino-acid sequence data will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under accession number Y10256.

Suppression of c-Myc-induced apoptosis by Ras signalling through PI(3)K and PKB

Andrea Kauffmann-Zeh*, Pablo Rodriguez-Viciana*, Eugen Ulrich*, Christopher Gilbert*, Paul Coffert†, Julian Downward* & Gerard Evan*

* Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK

† Department of Pulmonary Diseases, University Hospital Utrecht, Heidelberglaan 100, Utrecht 3584CX, The Netherlands

The viability of vertebrate cells depends on survival factors which activate signal transduction pathways that suppress apoptosis. Defects in anti-apoptotic signalling pathways are implicated in many pathologies including cancer, in which apoptosis induced by deregulated oncogenes must be forestalled for a tumour to become established. Phosphatidylinositol-3-kinase (PI(3)K) is involved in the intracellular signal transduction of many receptors and has been implicated in the transduction of survival signals in neuronal cells¹. We therefore examined the role of PI(3)K, its upstream effector Ras², and its putative downstream protein kinase effectors PKB/Akt^{3,4} and p70^{S6K} (ref. 5) in the modulation of apoptosis induced in fibroblasts by the oncoprotein c-Myc. Here we show that Ras activation of PI(3)K suppresses c-Myc-induced apoptosis through the activation of PKB/Akt but not p70^{S6K}. However, we also found that Ras is an effective promoter of apoptosis, through the Raf pathway. Thus Ras activates contradictory intracellular pathways that modulate cell viability. Induction of apoptosis by Ras may be an important factor in limiting the expansion of somatic cells that sustain oncogenic *ras* mutations.

In mesenchymal cells, induction of apoptosis by the pervasive oncogene *c-myc* is effectively suppressed by two principal cytokines, PDGF and IGF-I (ref. 6). PDGF and IGF-I, along with nerve growth factor (NGF) are also potent survival factors for cells of neuronal

origin. PI(3)K has been shown to be essential intermediary in NGF survival signalling¹, raising the possibility that PDGF and IGF-I also suppress apoptosis by activating PI(3)K through their cognate receptors^{7,8}. Indeed, we have previously observed that the well-characterized PI(3)K inhibitors wortmannin⁹ and LY294002 (ref. 10) both block IGF-I-mediated protection from apoptosis induced by *c-myc* in fibroblasts (data not shown), supporting a role for PI(3)K in fibroblast survival. The canonical PI(3)K comprises a regulatory p85 subunit and a catalytic p110 subunit that phosphorylates PtdIns(4,5)P₂ to create the second messenger PtdIns(3,4,5)P₃ (ref. 7). Putative downstream effectors of PI(3)K are the ribosomal protein kinase p70^{S6K} (ref. 5), the Rho family polypeptide Rac¹¹, and the serine/threonine protein kinase PKB/Akt^{3,4}.

To investigate further the role of PI(3)K in suppressing fibroblast apoptosis, we expressed a constitutively active 9E10 epitope-tagged

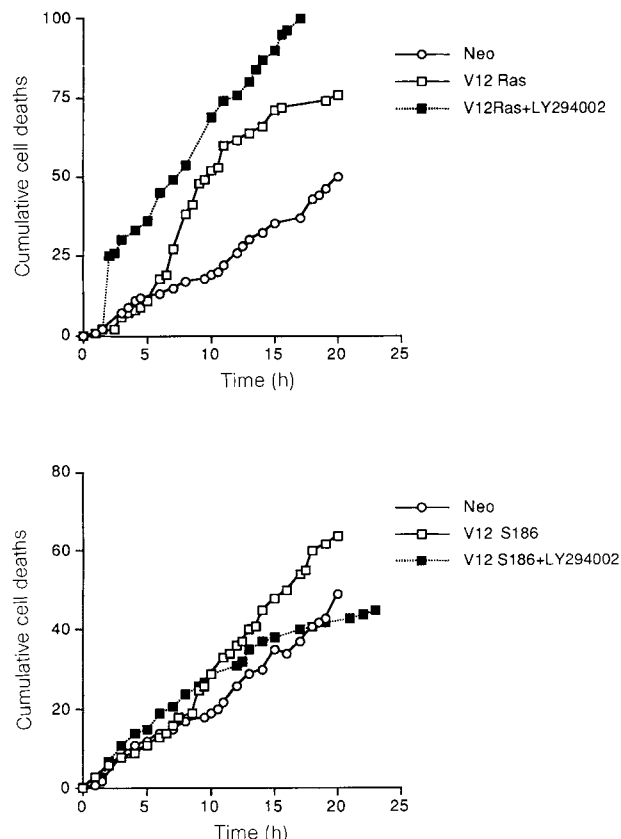


Figure 2 Effect of oncogenic Ras in the cell death survival ratio. Cells expressing the oncogenic V12 Ras, the farnesylation defective V12 S186 Ras, or vector alone (Neo) were serum starved for 48 hours and treated with 10 μ M LY294002 (LY) or DMSO carrier before activation of *c-Myc*-ERTM. Data plotted were extracted from time-lapse video-microscopy.

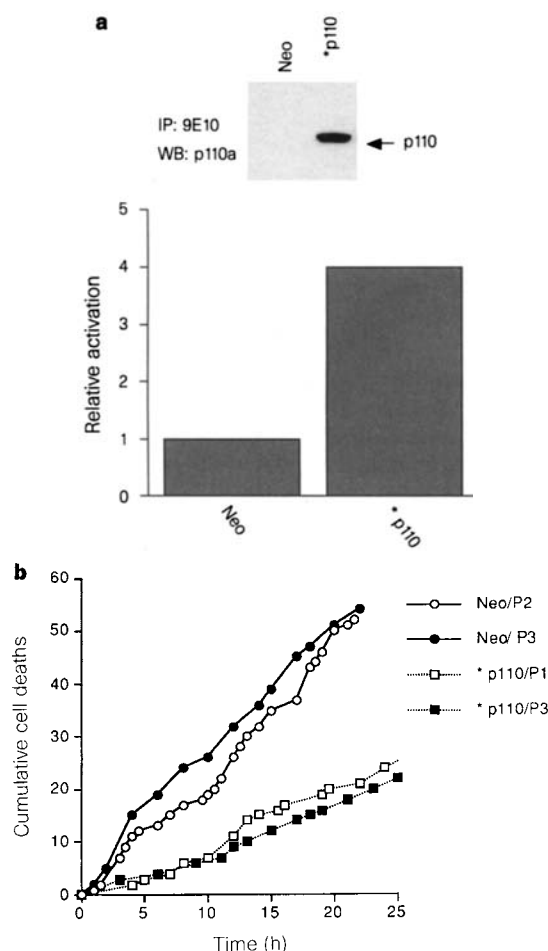
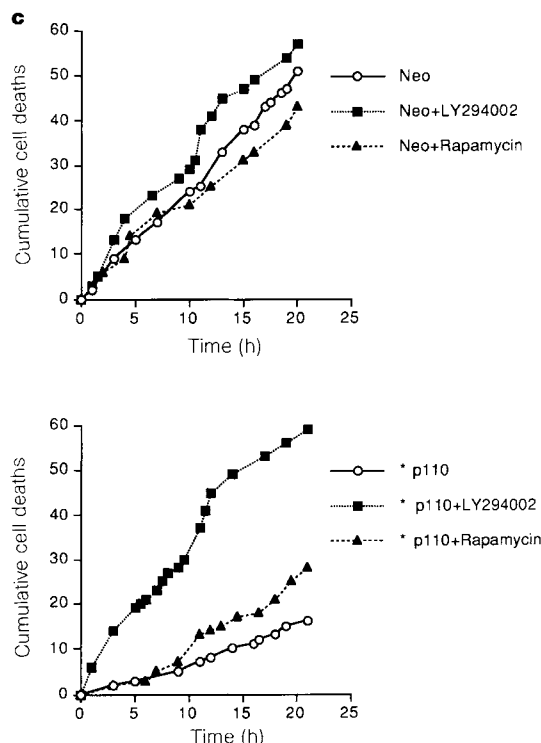


Figure 1 Constitutively active PI(3)K protects against *c-Myc*-induced apoptosis. **a**, Rat-1 cells expressing *c-Myc*-ERTM (R1MycMER) were stably transfected with 9E10-tagged p110 K227E PI(3)K mutant (*p110)¹² or vector alone (Neo). Expression of *p110 was confirmed by immunoprecipitation (IP) with 9E10 antibody followed by western blot (WB) analysis using a polyclonal antiserum raised against the C-terminal 14 amino acids of bovine p110 α (top). Approximately 48 h after transfection and after overnight withdrawal of serum, HA-PKB was immunoprecipitated with 12CA5 antibody and *in vitro* kinase assays were performed³ using histone H2B as substrate (bottom). **b**, Two distinct pools of each Neo and *p110 cells were growth arrested for 48 h in the absence of FCS, and *c-Myc*-ERTM was activated by addition of OHT. Cells were observed by time-lapse videomicroscopy. **c**, Serum-starved Neo (top) or *p110 (bottom) cells were treated with or without 10 μ M LY294002 (LY; Affinity Research) or 20 ng ml⁻¹ rapamycin (Rapa; Calbiochem-Novabiochem) and then assayed for suppression of *c-Myc*-induced apoptosis by time-lapse video-microscopy.



p110 227E mutant of PI(3)K (*p110)¹² in serum-deprived Rat-1 fibroblasts, and then induced apoptosis by activation of a 4-hydroxytamoxifen (OHT)-dependent conditional *c-myc* protein¹³. Expression of *p110 was confirmed by western immunoblotting and its activity verified by its ability to upregulate haemagglutinin epitope-tagged PKB/Akt (HA-PKB) kinase activity (Fig. 1a) and endogenous PKB/Akt (data not shown). Expression of *p110 delayed significantly the onset of apoptosis induced by *c-Myc* (Fig. 1b). The delay was similar in magnitude to that exerted by IGF-I (ref. 6) and was abrogated by either wortmanin (data not shown) or LY294002, but not by the p70^{S6K} inhibitor rapamycin¹⁴ (Fig. 1c). There was little or no effect on the death/survival ratio of control cells exposed to the same inhibitors (Fig. 1c). Thus PI(3)K activity suppresses apoptosis in response to oncogene action in fibroblasts.

The p110 K227E mutant is constitutively active because of a point mutation in its p21^{ras} (Ras) interaction domain that is thought to mimic the conformational change caused by binding to activated Ras¹². This suggests that PI(3)K-mediated suppression of apoptosis might be triggered by Ras activation, so we investigated whether the constitutively activated oncogenic V12 Ras mutant suppresses *c-Myc*-induced apoptosis in fibroblasts. However, expression of V12 Ras enhanced, rather than suppressed, apoptosis (Fig. 2, top); this enhancement was exacerbated by treatment of cells with the PI(3)K inhibitor LY294002 (Fig. 2). One possible explanation for this is that activated Ras triggers both a pro-apoptotic pathway and a separate PI(3)K-dependent anti-apoptotic pathway; in the absence of other factors, the former is dominant over the latter, such that the net effect of V12 Ras is to promote apoptosis.

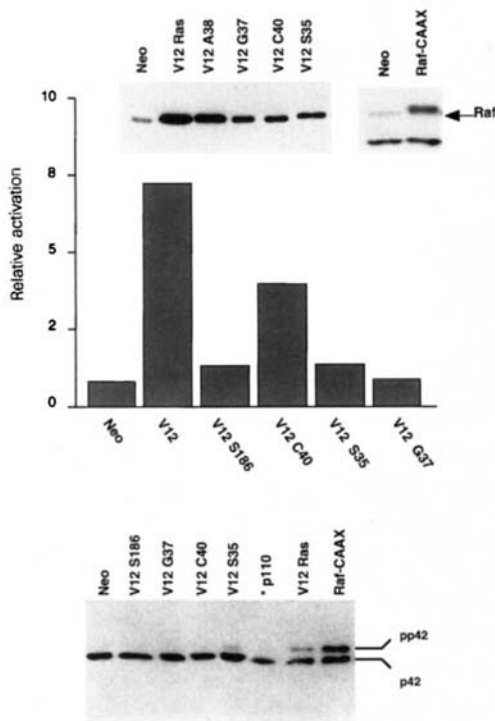


Figure 3 Expression of oncogenic Ras and Ras effector mutants in R1MycMER cells. Top, expression of V12 Ras, V12 S186 Ras, Raf-CAAX, various Ras effector mutants, or vector alone (Neo) was confirmed by western immunoblotting using anti-Ras (Oncogene Science) and anti-Raf (Santa Cruz) antibodies (see inset). These cells were transfected with HA-PKB and its activity determined³ following immunoprecipitation with 12CA5 antibody (see Methods). Bottom, R1MycMER cells expressing V12 Ras, V12 S186 Ras, various Ras effector mutants, the constitutively activated Raf-CAAX mutant²⁷, or vector only (Neo) were analysed for phosphorylation of ERK2 by mobility shift of the p42 MAP kinase to the phosphorylated pp42 form.

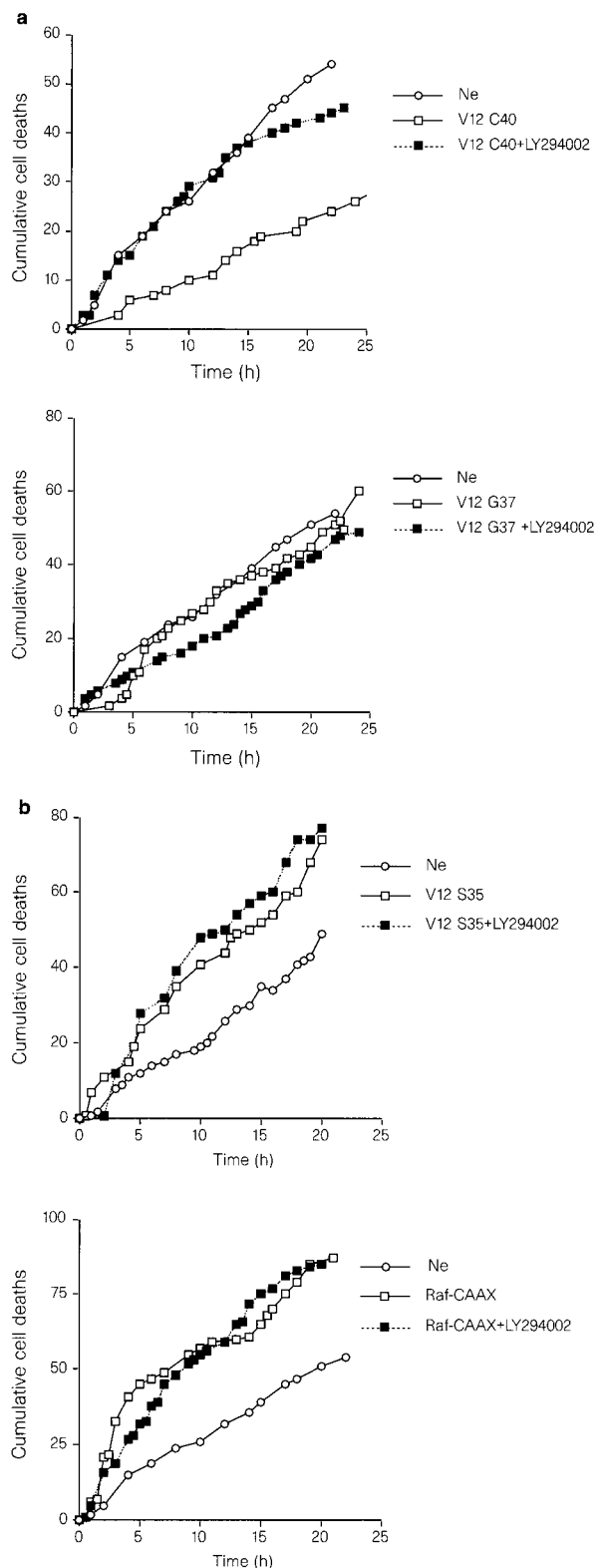


Figure 4 Evidence for a dual effect of Ras on *c-Myc*-induced apoptosis. **a**, Serum-starved R1MycMER cells expressing vector alone (Neo) or the effector mutants V12 C40 Ras and V12 G37 Ras were induced to undergo apoptosis by activation of *c-Myc* in the presence of 10 μ M LY294002 (LY) or DMSO carrier. Cells were monitored by time-lapse videomicroscopy. **b**, Growth-arrested R1MycMER cells expressing Raf-activating Ras effector mutant V12 S35, constitutively activated Raf-CAAX, and vector only (Neo), were treated with 10 μ M LY294002 (LY) or the carrier DMSO before activation of *c-Myc*-ERTM. Cells were analysed by time-lapse video-microscopy.

Active, GTP-bound Ras transduces signals through multiple intracellular targets. These include, amongst others, the p110 catalytic subunit of PI(3)K¹², Raf (at the apex of the MAP kinase pathway¹⁵⁻¹⁷), and Ral.GDS, the exchange factor for Ral.GTPases⁸⁻²⁰. We therefore investigated the individual contributions made by each of these effectors to cell survival. We used partial loss-of-function mutants located in the Ras effector loop that each activates only one of the downstream pathways mentioned above (P.R.-V., manuscript submitted). Each Ras effector mutant resided in a constitutively activated V12 Ras background. Thus oncogenic V12 Ras interacts strongly and constitutively with, and activates, all known Ras effectors; in contrast, the farnesylation-defective V12 S186 double mutant activates none^{21,22}. The partial loss-of-function mutant V12 S35 (ref. 23) interacts with Raf but not with PI(3)K or Ral.GDS; the V12 C40 mutant²⁴ interacts with PI(3)K p110 α but not with Raf or Ral.GDS; and V12 G37 Ras²³ interacts with Ral.GDS but not with Raf or PI(3)K (P.R.-V., manuscript submitted). These various V12 Ras effector mutants were each expressed in Rat-1 fibroblasts (Fig. 3) and assessed for their abilities to suppress c-Myc-induced apoptosis.

Activation of the Ral.GDS pathway by expression of the V12 G37 Ras mutant had no detectable effect on c-Myc-induced apoptosis, excluding this pathway from the modulation of apoptosis. The function of Ral is currently unknown, although a role in the regulation of phospholipase D²⁵ and Rho family proteins^{26,27} has been suggested. As expected, expression of the V12 S186 total loss-of-function mutant also had no effect on c-Myc-induced apoptosis (Fig. 2). Unlike oncogenic V12 Ras, which strongly activates both MAP kinase (ERK2) and the PI(3)K target HA-PKB, V12 S186 activated neither in the absence of serum (Fig. 3). Expression of V12 C40 Ras protected cells from c-Myc-induced death as effectively as a constitutively active K227E p110 PI(3)K mutant (Fig. 4a), and this protection was abolished by treatment of cells with the PI(3)K inhibitor LY294002. HA-PKB activity in cells expressing V12 C40 Ras was higher than in control cells but, as expected, was lower than that observed in cells expressing oncogenic V12 Ras, which is a stronger activator of PI(3)K (P.R.-V., manuscript submitted). Predictably, V12 C40 did not induce phosphorylation of MAP kinase (Fig. 3). To investigate the possible role of the MAP kinase pathway in modulating apoptosis, we examined the effects of the V12 S35 effector mutant and of a constitutively activated Raf-CAAX

mutant²⁸ (Fig. 4b). Both Raf-CAAX and V12 S35 Ras mutants increased c-Myc-induced apoptosis (Fig. 4b), and this was unaffected by inhibition of PI(3)K with LY294002. Cells expressing either of these mutants exhibited activation of MAP kinase but no increase in HA-PKB activity (Fig. 3 and data not shown). Taken together, these data suggest that Ras-mediated activation of the Raf pathway promotes apoptosis in fibroblasts, whereas Ras-mediated activation of the PI(3)K pathway suppresses it.

We next considered possible downstream effectors of PI(3)K that are involved in suppressing apoptosis. We could exclude one such effector, p70^{S6K}, because the macrolide rapamycin, which induces dephosphorylation and concomitant inactivation of p70^{S6K} (ref. 14), had no detectable inhibitory effect on PI(3)K-mediated suppression of apoptosis (Fig. 1c). Another candidate downstream effector of PI(3)K is the serine/threonine protein kinase PKB/Akt, which is activated by both *p110 PI(3)K and the PI(3)K-specific mutant Ras V12 C40 (Figs 1a and 3). To examine the potential role of PKB/Akt kinase, we transiently expressed HA-PKB, a constitutively activated *gag*-PKB/Akt fusion protein that is similar to the product of the retrovirus *v-Akt* oncogene (*gag*-PKB), or a catalytically inactive *gag*-PKB/Akt in which the critical lysine of the ATP-binding site has been replaced by alanine (HA-PKB.K>A); we then examined the sensitivity of the transfected cells to c-Myc-induced apoptosis. Compared with cells harbouring the pSG5-*gag* control vector, cells expressing *gag*-PKB were substantially protected from apoptosis (Fig. 5). Cells expressing HA-PKB exhibited limited resistance to apoptosis, whereas cells expressing the kinase-defective *gag*-PKB.K>A mutant behaved similarly to control cells. These effects were more evident when cells were exposed to PDGF, which is known to activate PKB/Akt kinase activity^{3,4} (Fig. 5). The protection conferred by PDGF treatment and the partial protection observed after transfection with HA-PKB were both abolished when cells were exposed to the PI(3)K inhibitor LY294002. In contrast, the suppression of apoptosis afforded by cells expressing *gag*-PKB was not affected by treatment with LY294002 (Fig. 5). It is therefore likely that PI(3)K-mediated activation of the PKB/Akt kinase is a significant component of the anti-apoptotic signalling pathway.

Although much is known about the basal execution machinery of apoptosis, in particular the ICE-like cysteine proteases²⁹, almost nothing is known about how this machinery is modulated by anti-apoptotic signalling pathways. Our data trace a putative pathway through Ras and PI(3)K to PKB/Akt. It now becomes important to search for downstream targets of PKB/Akt that might be involved in modulating the apoptotic machinery. At present, the only known potential downstream target of PKB/Akt is glycogen synthase kinase-3 (ref. 30), to which no role in apoptosis has yet been ascribed.

Our findings underscore the pleiotropic nature of intracellular signalling. Signals emanating from GTP-Ras trigger a plethora of potential biological outcomes, some of which, as in the case of promotion (Raf pathway) and suppression (PI(3)K/PKB pathway) of apoptosis, may be contradictory. The net outcome of Ras activation is presumably dictated by downstream interactions that independently potentiate or mitigate each effector pathway in response to other signals. An analogy might be made with c-Myc, which can induce both proliferation or apoptosis but whose net effect depends upon the status of anti-apoptotic signals within the cell. The pro-apoptotic proclivities of c-Myc and Ras may thus serve to limit the expansion and carcinogenic progression of cells in which they become oncogenically activated.

Methods

Generation of cell lines. Constitutively active PI(3)K, Raf-CAAX, V12 Ras, V12 S186 Ras, and the various Ras effector mutants were cloned into the Moloney murine leukaemia-virus-based retroviral vector pLXSN. These constructs were stably transfected into the ecotropic packaging cell line GP + E cells using Lipofectamine (Gibco/BRL). Virus-containing supernatants

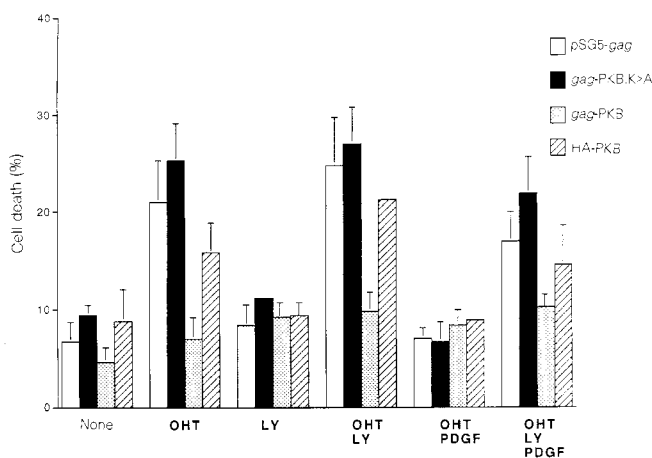


Figure 5 The putative PI(3)K target PKB/Akt confers protection against c-Myc-induced cell death. Sub-confluent R1MycMER cultures were transfected with HA-PKB, *gag*-PKB.K>A (kinase dead), *gag*-PKB (constitutively active), and control vector (pSG5-*gag*)³ using Lipofectamine (Gibco/BRL). Approximately 48 h later, triplicate dishes were serum starved for 18–24 h before treatment with or without PDGF (50 ng ml⁻¹) and 100 nM OHT in the presence of 10 μ M LY294002 or DMSO carrier. Both floating and attached cells were recovered after 15–18 h and assessed for DNA fragmentation by flow cytometry.

were used to infect Rat-1 cells expressing the conditional c-Myc allele c-Myc-ERTM (R1MycMER) in the presence of Polybrene ($4 \mu\text{g ml}^{-1}$).

C-Myc activation. c-Myc was activated by the addition of 100 nM of 4-hydroxytamoxifen to cells that have been serum deprived for 25–28 h (ref. 13).

PKB/Akt assays. *In vitro* kinase assays were performed on immunoprecipitated PKB/Akt proteins³ using histone H2B as substrate. To this end, PKB/Akt constructs were transfected into R1MycMER using Lipofectamine (Gibco/BRL) and, approximately 48 h later, cells were serum deprived for 24 h before immunoprecipitation with 12CA5 antibody. Endogenous PKB/Akt kinase activation was analysed in immunoprecipitates obtained from serum-deprived cells by using the anti-Rac-CT antibody (Upstate Biotechnology).

ERK2 assays. ERK2 phosphorylation was determined by mobility shift of p42 MAP kinase to its phosphorylated pp42 form. To this end, cells were serum deprived for 48 h, total cell lysates electrophoresed in 15% low bis-acrylamide gels, and western immunoblotting performed using anti-ERK2 (Upstate Biotechnology).

Cell death analysis. Flow cytometric analysis for apoptotic DNA fragmentation was performed using Apotag (Oncor Appligene) on cells fixed with 1% paraformaldehyde. Time-lapse video-microscopic analysis was performed as described⁴.

Received 21 August 1996; accepted 6 January 1997.

1. Yao, R. & Cooper, G. M. Requirement for phosphatidylinositol-3 kinase in the prevention of apoptosis by nerve growth factor. *Science* **267**, 2003–2006 (1995).
2. Rodriguez Viciano, P. *et al.* Phosphatidylinositol-3-OH kinase as a direct target of Ras. *Nature* **370**, 527–532 (1994).
3. Burgering, B. M. & Coffey, P. J. Protein kinase B (c-Akt) in phosphatidylinositol-3-OH kinase signal transduction. *Nature* **376**, 599–602 (1995).
4. Franke, T. F. *et al.* The protein kinase encoded by the Akt proto-oncogene is a target of the PDGF-activated phosphatidylinositol 3-kinase. *Cell* **81**, 727–736 (1995).
5. Chung, J., Grammer, T. C., Lemon, K. P., Kazlauskas, A. & Blenis, J. PDGF- and insulin-dependent pp70^{S6K} activation mediated by phosphatidylinositol-3-OH kinase. *Nature* **370**, 71–75 (1994).
6. Harrington, E., Fanidi, A., Bennett, M. & Evan, G. Modulation of Myc-induced apoptosis by specific cytokines. *EMBP J.* **13**, 3286–3295 (1994).
7. Kapeller, R. & Cantley, L. C. Phosphatidylinositol 3-kinase. *Bioessays* **16**, 565–576 (1994).
8. Seely, B. L. *et al.* Localization of the insulin-like growth factor I receptor binding sites for the SH2 domain proteins p85, Syp, and GTPase activating protein. *J. Biol. Chem.* **270**, 19151–19157 (1995).
9. Ui, M., Okada, T., Hazeki, K. & Hazeki, O. Wortmannin as a unique probe for an intracellular signalling protein, phosphoinositide 3-kinase. *Trends Biochem. Sci.* **20**, 303–307 (1995).
10. Vlahos, C. J., Matter, W. F., Hui, K. Y. & Brown, R. F. A specific inhibitor of phosphatidylinositol 3-kinase, 2-(4-morpholinyl)-8-phenyl-4H-1-benzopyran-4-one (LY294002). *J. Biol. Chem.* **269**, 5241–5248 (1994).
11. Hawkins, P. T. *et al.* PDGF stimulates an increase in GTP-Rac via activation of phosphoinositide 3-kinase. *Curr. Biol.* **5**, 393–403 (1995).
12. Rodriguez-Viciano, P., Vanhaesebroeck, B., Waterfield, M. D. & Downward, J. Activation of phosphoinositide 3-kinase by interaction with Ras and by point mutation. *EMBO J.* **15**, 2442–2451 (1996).
13. Littlewood, T., Hancock, D., Danielian, P., Parker, M. & Evan, G. A modified oestrogen receptor ligand-binding domain as an improved switch for the regulation of heterologous proteins. *Nucleic Acids Res.* **23**, 1686–1690 (1995).
14. Chung, J., Kuo, C. J., Crabtree, G. R. & Blenis, J. Rapamycin-FKBP specifically blocks growth-dependent activation of and signalling by the 70 and 86 protein kinases. *Cell* **69**, 1227–1236 (1992).
15. Warne, P. H., Viciano, P. R. & Downward, J. Direct interaction of Ras and the amino-terminal region of Raf-1 *in vitro*. *Nature* **364**, 352–355 (1993).
16. Zhang, X. F. *et al.* Normal and oncogenic p21^{ras} proteins bind to the amino-terminal regulatory domain of c-Raf-1. *Nature* **364**, 308–313 (1993).
17. Vojtek, A. B., Hollenberg, S. M. & Cooper, J. A. Mammalian Ras interacts directly with the serine/threonine kinase Raf. *Cell* **74**, 205–214 (1993).
18. Albright, C. E., Giddings, B. W., Liu, J., Vito, M. & Weinberg, R. A. Characterization of a guanine nucleotide dissociation stimulator for a ras-related GTPase. *EMBO J.* **12**, 339–347 (1993).
19. Kikuchi, A., Demo, S. D., Ye, Z. H., Chen, Y. W. & Williams, L. T. ralGDS family members interact with the effector loop of ras p21. *Mol. Cell. Biol.* **14**, 7483–7491 (1994).
20. Spaargaren, M. & Bischoff, J. R. Identification of the guanine nucleotide dissociation stimulator for Ras as a putative effector molecule of R-ras, H-ras, and Rap. *Proc. Natl. Acad. Sci. USA* **91**, 12609–12613 (1994).
21. Leever, S. J. & Marshall, C. J. Activation of extracellular signal-regulated kinase, ERK2, by p21ras oncoprotein. *EMBO J.* **11**, 569–574 (1992).
22. Cadwallader, K. A., Paterson, H., Macdonald, S. G. & Hancock, J. F. N-terminally myristoylated Ras proteins require palmitoylation or a polybasic domain for plasma membrane localization. *Mol. Cell. Biol.* **14**, 4722–4730 (1994).
23. White, M. A. *et al.* Multiple Ras functions can contribute to mammalian cell transformation. *Cell* **80**, 533–541 (1995).
24. Joneson, T., White, M. A., Wigler, M. H. & Bar Sagi, D. Stimulation of membrane ruffling and MAP kinase activation by distinct effectors of RAS. *Science* **271**, 810–812 (1996).
25. Jiang, H. *et al.* Involvement of Ral GTPase in v-Src-induced phospholipase D activation. *Nature* **378**, 409–412 (1995).
26. Julien Flores, V. *et al.* Bridging Ral GTPase to Rho pathways. RLIP76, a Ral effector with CDC42/Rac GTPase-activating protein activity. *J. Biol. Chem.* **270**, 22473–22477 (1995).
27. Park, S. H. & Weinberg, R. A. A putative effector of Ral has homology to Rho/Rac GTPase activating proteins. *Oncogene* **11**, 2349–2355 (1995).
28. Leever, S. J., Paterson, H. F. & Marshall, C. J. Requirement for Ras in Raf activation is overcome by targeting Raf to the plasma membrane. *Nature* **369**, 411–414 (1994).
29. Takahashi, A. & Earnshaw, W. ICE-related proteases in apoptosis. *Curr. Opin. Genet. Dev.* **6**, 50–55 (1996).
30. Cross, D. A., Alessi, D. R., Cohen, P. & Andjelkovich, M. & Hemmings, B. A. Inhibition of glycogen synthase kinase-3 by insulin mediated by protein kinase B. *Nature* **378**, 785–789 (1995).

Acknowledgements. We thank D. Lyon for help with the time-lapse video-microscopy, and D. Davies for the FACS analysis. A.K.-Z. is a fellow of the Kay Kendall Leukaemia Fund. E. Ulrich is a fellow of the F.W.F. (Schrodinger Stipendium). We thank the Bishop Stopford School, Kettering, UK, for a generous donation in memory of Elliot Smith.

Correspondence and requests for materials should be addressed to G.E. (e-mail: g.evan@icr.icnet.uk).

The nuclear receptor homologue Ftz-F1 and the homeodomain protein Ftz are mutually dependent cofactors

Antoine Guichet*, John W. R. Copeland†, Miklós Erdélyi*||, Daniela Hlousek†, Péter Závorszky*, Jacqueline Ho‡, Susan Brown§, Anthony Percival-Smith‡, Henry M. Krause† & Anne Ephrussi*

* Developmental Biology Programme, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

† Banting and Best Department of Medical Research and Department of Molecular and Medical Genetics, University of Toronto, Charles H. Best Institute, 112 College Street, Toronto, Ontario, Canada, M5G 1L6

‡ Department of Zoology, University of Western Ontario, London, Ontario, Canada, N6A 5B7

§ Division of Biology, Ackert Hall, Kansas State University, Manhattan, Kansas 66506, USA

Nuclear hormone receptors and homeodomain proteins are two classes of transcription factor that regulate major developmental processes. Both depend on interactions with other proteins for specificity and activity. The *Drosophila* gene *fushi tarazu* (*ftz*), which encodes a homeodomain protein¹ (Ftz), is required zygotically for the formation of alternate segments in the developing embryo². Here we show that the orphan nuclear receptor α Ftz-F1 (ref. 3), which is deposited in the egg during oogenesis⁴, is an obligatory cofactor for Ftz. The two proteins interact specifically and directly, both *in vitro* and *in vivo*, through a conserved domain in the Ftz polypeptide. This interaction suggests that other nuclear receptor/homeodomain protein interactions may be important and common in developing organisms.

In a screen for new maternal mutations affecting anteroposterior polarity, we identified two recessive mutations causing a pair-rule phenotype that map near the *Ftz-F1* gene³ (Fig. 1). Embryos derived from homozygous mutant females lack alternate denticle belts, normally found in segments T2, A1, A3, A5 and A7 (Fig. 1b). Closer examination of the cuticles showed that these deletions are parasegmental in nature (data not shown), as are those of *ftz* (ref. 2) (Fig. 1c).

The *Ftz-F1* gene encodes two protein isoforms, α and β . The α isoform is maternally expressed and evenly distributed in the early embryo⁵. In contrast, the β isoform is zygotically expressed during late embryonic and pupal stages, at which time it is thought to play a role in ecdysone-induced gene expression^{4,6}. The two alleles isolated in this study, *Ftz-F1*²⁰⁹ and *Ftz-F1*²⁸², produce indistinguishable mutant phenotypes. Two lines of evidence indicate that both alleles specifically compromise maternal expression of the *Ftz-F1* gene. First, northern blots show that α Ftz-F1 transcripts are not detected in *Ftz-F1*²⁸² mutant females, and that a truncated transcript is detected in *Ftz-F1*²⁰⁹ mutant females (Fig. 1d). Second, *Ftz-F1*²⁰⁹ mutant embryos are rescued by expression of the α Ftz-F1 transcript under control of the *hsp70* heat-shock promoter (Fig. 1e–h). This experiment was possible because the maternal product is not spatially localized⁵. The finding that, at blastoderm stage, the maternal α Ftz-F1 gene product is uniformly distributed (data not shown), yet its mutation causes a pair-rule phenotype, indicates a limited requirement for α Ftz-F1 in alternate parasegments. Further, it suggests that α Ftz-F1 must interact with spatially localized factors.

|| Present address: Biological Research Center, Hungarian Academy of Sciences, Institute of Genetics, H-6701 Szeged POB 521, Hungary

α Ftz-F1 was first identified as a protein that binds the *ftz* promoter⁷. It was subsequently shown that α Ftz-F1 binding sites are present in two regulatory regions of the *ftz* promoter, the proximal zebra element⁷ and the upstream enhancer element⁸, and that integrity of these sites is important for proper expression of minimized *ftz-lacZ* reporter genes^{5,7-9}. It therefore seemed likely that the α Ftz-F1 phenotype is a consequence of reduced *ftz* expression. Surprisingly, patterns of *ftz* mRNA and protein expression in our α Ftz-F1 mutants are indistinguishable from wild type (Fig. 2a-d). We therefore tested whether other *ftz*-dependent parasegmental markers are correctly expressed. Ftz is required for proper expression of two segment-polarity genes, *engrailed* (*en*) and *wingless* (*wg*), which are each expressed in 14 stripes^{10,11} (Fig. 2g, j). Even-numbered *en* stripes are positively regulated by Ftz (ref. 12), whereas even-numbered *wg* stripes are negatively regulated by Ftz (ref. 13). In *Ftz-F1* mutant embryos, as in *ftz* mutant embryos (Fig. 2i, l), Ftz-dependent *en* stripes fail to be expressed (Fig. 2f, h), and *wg* stripes expand (Fig. 2k). Thus α Ftz-F1 is required for all Ftz activities tested except that for which it was first identified: regulation of the *ftz* promoter.

The target gene specificity of homeodomain proteins is critically dependent on extrinsic factors, such as interactions with other proteins. Indeed, Ftz can regulate many target genes in the absence of its DNA-binding homeodomain^{14,15}. It has been suggested that this activity is mediated by DNA-binding cofactors^{14,15}. Our results suggest the possibility that α Ftz-F1 may function as one of these cofactors. To examine this possibility, we tested for a direct interaction between Ftz and α Ftz-F1 *in vitro*. Nearly all of the α Ftz-F1 protein expressed in a reticulocyte lysate system is specifically retained on Ftz micro-affinity columns with no binding to control columns (Fig. 3a). Retention of α Ftz-F1 on the Ftz affinity columns

is consistently as good as or better than that observed with the previously demonstrated¹⁵ Ftz-interacting protein, Paired (Prd).

Further analysis of the α Ftz-F1/Ftz interaction by far western blotting (Fig. 3b) confirmed that the interaction is direct and does not require the Ftz homeodomain. Indeed, the N-terminal third of Ftz is sufficient for a strong interaction (Fig. 3c). By using additional deletion constructs, the interaction domain was narrowed down to residues 101-150 (Fig. 3c). Two regions of the Ftz polypeptide are highly conserved in dipteran Ftz homologues: the homeodomain, and residues within the Ftz Δ 101-150 deletion. In *Drosophila hydei*, residues corresponding to amino acids 100-133 of Ftz are 100% conserved¹⁶. In the more distantly related flour beetle (*Tribolium castaneum*) homologue (TcFtz), the only homology, other than the homeodomain, is within this N-terminal region. This homology is more limited, however, consisting of a central LRALLT motif flanked on either side by prolines, adjacent to which are acidic residues on the N-terminal side and basic residues on the C-terminal side¹⁷. Despite this more limited homology, TcFtz interacts strongly with α Ftz-F1 on far western blots, and the N-terminal region is sufficient for this interaction (Fig. 3c). Taken together, these data suggest that the LRALLT domain is the primary contact point for Ftz-F1, and that this domain is sufficiently important to have warranted conservation over >300,000 yr (ref. 17).

It is thought that *ftz* is a rapidly evolving homeobox gene that was first required in insects for development of the central nervous system, and only more recently for segmentation^{17,18}. The *Tribolium* *ftz* homologue may represent an evolutionary *ftz* intermediate, as it is segmentally expressed but has no obvious segmental phenotype¹⁹. The ability of TcFtz to bind α Ftz-F1 suggests that *Tribolium* may also possess a *Ftz-F1* homologue, and that this gene may hold the key to the evolving role of Ftz as a pair-rule protein.

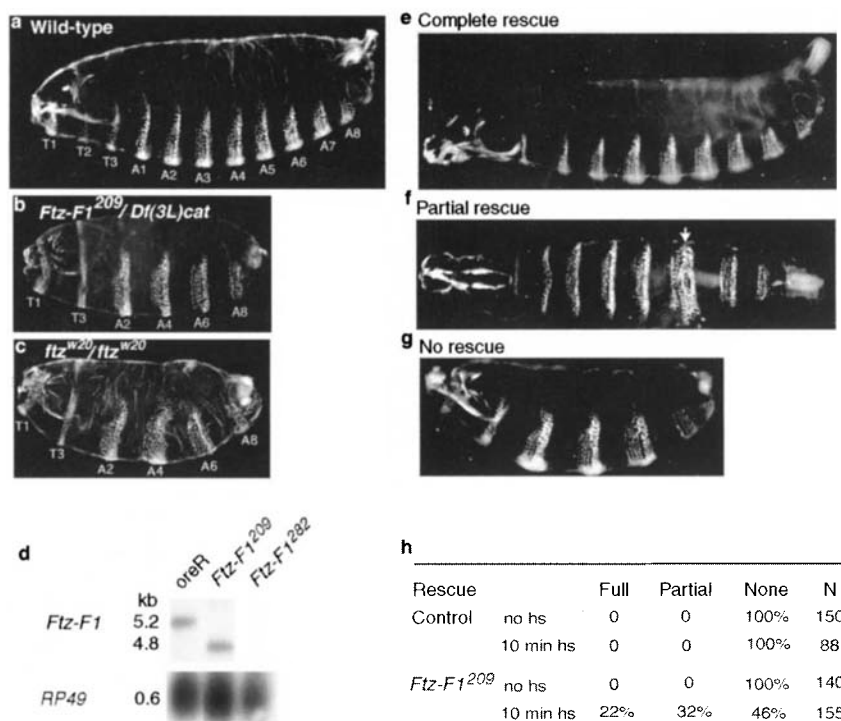
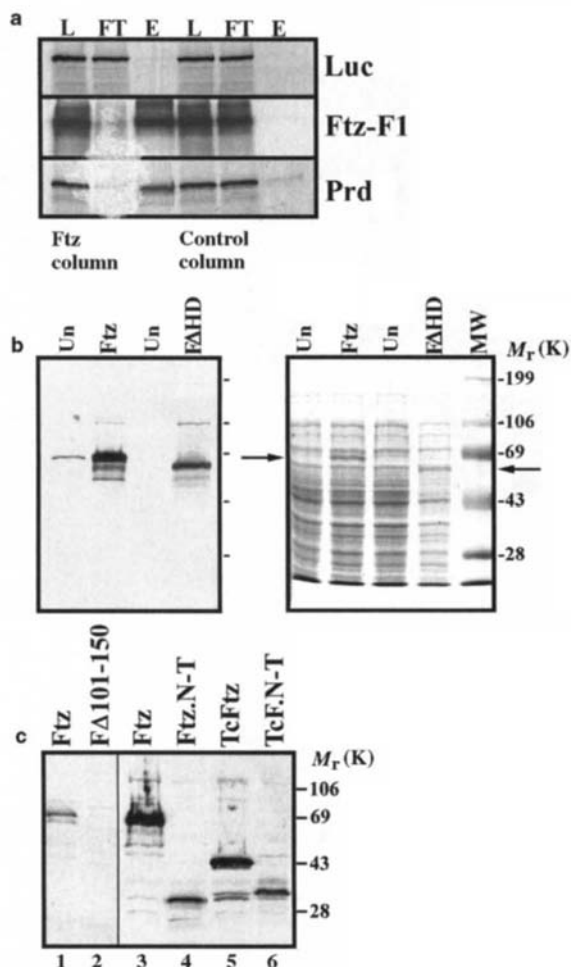
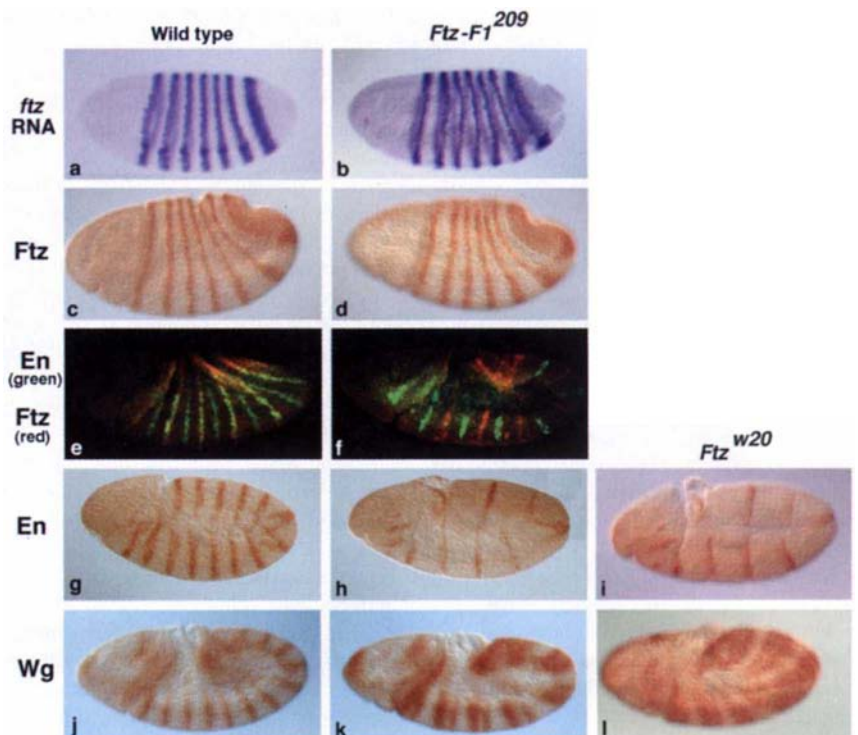


Figure 1 Mutation of the *ftz-F1* α isoform leads to a maternal-effect, pair-rule phenotype that is indistinguishable from that of *ftz*. **a**, Wild-type (*w*¹¹⁸) embryos. *ftz-F1*²⁰⁹/*Df*(3L)*cat* (**b**) and *ftz*^{w20}/*ftz*^{w20} (**c**) embryos have indistinguishable cuticular phenotypes in which segments T2, A1, A3, A5, A7 are deleted. **d**, Northern blot containing RNA isolated from adult *oreR* (wild type), *Ftz-F1*²⁰⁹, and *Ftz-F1*²⁸² homozygous females. Hybridization with a 5' probe specific for the α isoform reveals a single 5.2-kb transcript in wild-type females. This transcript corre-

sponds to the α Ftz-F1 mRNA⁴. In *Ftz-F1*²⁰⁹, a shorter transcript (4.8 kb) is present, whereas in *Ftz-F1*²⁸² no transcript is detected. A probe specific for ribosomal protein 49 (RP49) serves as loading control. **e-h**, Zygotic rescue of the segmentation defects of *Ftz-F1*²⁰⁹ by a heat-shock (*hs*) transgene *hs* α Ftz-F1. **f**, The white arrow in **f** indicates the remaining fused segments A5-A6 in the partially rescued embryo.

Figure 2 *Ftz-F1* affects *en* and *wg*, but not *ftz*, expression. **a, c, e, g, i, j**, Wild-type embryos. **b, d, f, h, k**, Embryos from *Ftz-F1²⁰⁹/Df(3L)cat* females, referred to as *Ftz-F1²⁰⁹* embryos. **i, l**, Homozygous *ftz^{w20}* embryos. **a, b**, *ftz* mRNA expression in stage-5 embryos. The pattern of expression of *ftz* mRNA in *Ftz-F1²⁰⁹* embryos (**b**), is indistinguishable from that in wild-type embryos (**a**). **c, d**, *Ftz* protein expression in stage-6 embryos. As is the case for *ftz* mRNA, the pattern of *Ftz* protein expression in *Ftz-F1²⁰⁹* embryos (**d**) is indistinguishable from that of wild-type embryos (**c**). **e, f**, Double-labelling revealing *En* protein (green) and *Ftz* protein (red) expression patterns. In *Ftz-F1²⁰⁹* embryos (**f**), *En* is absent from the stripes expressing *Ftz*, indicating that correctly expressed *Ftz* protein cannot activate *en* in the absence of *Ftz-F1*. **g-i**, *En* protein expression in stage-9 embryos. In wild-type embryos (**g**), *En* is expressed in 14 stripes, whereas in *Ftz-F1²⁰⁹* (**h**) and in *ftz^{w20}* embryos (**i**), *En* is expressed in 7 stripes. **j-l**, *Wg* protein expression in stage-8 embryos. In wild-type embryos (**j**), *Wg* is expressed as 14 stripes, but in *Ftz-F1²⁰⁹* embryos (**k**), *Wg* is expressed as 7 broad stripes, as in *ftz^{w20}* embryos (**l**). The late expression pattern of *runt* is also affected in *Ftz-F1*, as in *ftz^{w20}* embryos (data not shown). Expression of *even-skipped* and *hairy* is not affected in *Ftz-F1²⁰⁹* mutants (data not shown).



To test whether α Ftz-F1 and Ftz interact directly *in vivo*, we tested the ability of an ectopically expressed Ftz polypeptide, missing residues 101–150, to regulate α Ftz-F1-dependent target genes. When expressed under control of a heat-shock promoter, ubiquitous Ftz expression in blastoderm embryos represses alternate *wg* stripes, broadens alternate *en* stripes, and broadens endogenous *ftz* stripes²⁰ (Fig. 4). This results in an ‘anti-*ftz*’ pair-rule phenotype in which *ftz*-independent parasegments are deleted²¹ (Fig. 4n). These effects do not require the Ftz homeodomain, nor do they depend on the endogenous *ftz* gene^{14,15}.

Deletion of amino acids 101–150 disrupts all but one of the Ftz activities described above (Fig. 4). Ftz Δ 101–150 cannot repress *wg*,

Figure 3 Ftz and α Ftz-F1 interact directly. **a**, Ftz affinity chromatography. Reticulocyte lysate-translated Luciferase (Luc), α Ftz-F1 and Prd proteins were passed over Ftz (left) or control (right) affinity columns. Equivalent portions of the ³⁵S-labelled load (L), flow-through (FT), and eluate (E) fractions were subjected to SDS-PAGE and autoradiography. Luciferase, used as a negative control, flows through both the Ftz and control columns. In contrast, α Ftz-F1 is specifically retained on the Ftz affinity column and quantitatively recovered in the eluate. Prd, a previously identified Ftz-interacting¹⁵ protein, also binds specifically to the Ftz affinity column, but with lower efficiency. All other proteins within the reticulocyte lysate flowed through the column (ref. 15 and data not shown). **b**, Far western analysis. Autoradiogram (left) and corresponding Coomassie blue stain (right) show uninduced (Un) and induced bacterial lysates containing Ftz or Ftz Δ HD¹⁴ subjected to SDS-PAGE. Left, proteins were transferred from the gel to nitrocellulose. Right, gel was stained with Coomassie blue to reveal total proteins. α Ftz-F1 bound specifically to the induced Ftz and Ftz Δ HD proteins marked with arrows on the Coomassie-stained gel. MW, molecular weight markers. **c**, Far western with deleted Ftz polypeptides probed with ³⁵S-labelled α Ftz-F1. Lanes 1 and 2 contain full-length and N-terminally deleted (residues 101–150) Ftz polypeptides expressed *in vitro*. Lanes 3–6 contain *Drosophila* and *Tribolium* Ftz polypeptides expressed in bacteria: lane 3, full-length Ftz; lane 4, Ftz 1–171; lane 5, TcFtz; lane 6, TcFtz 1–197. All Ftz proteins except Ftz Δ 101–150 bound well to the α Ftz-F1 probe.

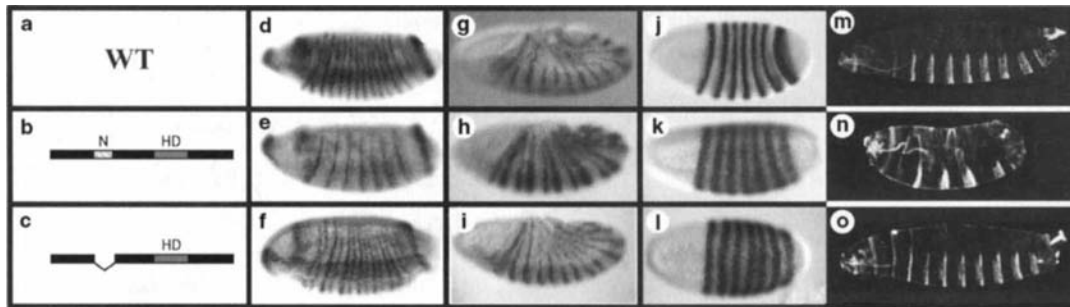


Figure 4 Deletion of the α Ftz-F1 interaction domain disrupts α Ftz-F1-dependent Ftz activities. The left hand column (a-c) indicates Ftz constructs expressed upon heat shock and responsible for the corresponding patterns in the panels to the right. **a**, No construct expressed (WT); **b**, full-length Ftz; **c**, Ftz Δ 101-150. Adjacent columns, from left to right, show resulting *wg* (d-f), *en* (g-i) and *ftz* (j-l) expression

activate *en* or generate an anti-*ftz* cuticular phenotype, although it is still capable of broadening endogenous *ftz* stripes. Similar results were obtained by expressing a full-length Ftz polypeptide in an α Ftz-F1 mutant background (data not shown). Thus removal of the α Ftz-F1 interaction domain from the Ftz polypeptide results in the same loss of Ftz activities as removal of α Ftz-F1.

Previous studies indicated that Ftz-mediated repression of *wg* requires the pair-rule protein Prd, and that this requirement involves a direct interaction between the two proteins¹⁵. The expanded pattern of *wg* expression in *Ftz-F1* mutants is not due to a defect in Prd expression, as the pattern of expression of Prd in *Ftz-F1* mutant embryos is indistinguishable from the wild-type pattern (data not shown). Thus Ftz-mediated repression of *wg* appears to require both Prd and α Ftz-F1. This interaction could involve either simultaneous or competitive interactions amongst the three proteins, as Prd also contacts residues 101-150 of Ftz (ref. 15). Unlike Ftz-F1, however, Prd requires additional contact points on the Ftz polypeptide for a strong interaction (J.W.R.C. and H.M.K., manuscript in preparation).

Prd may be a cofactor of Ftz or Ftz-F1 that is required for target genes that are repressed by Ftz, because Prd is required for Ftz-dependent *wg* repression, but not for Ftz-dependent activation of *en* (ref. 12) or for *ftz* auto-regulation²². For the latter two genes, no Ftz cofactors had previously been identified. In the case of *en* regulation, recent studies³¹ support our findings that Ftz and α Ftz-F1 are both required, and suggest a likely *cis*-acting target element. Binding sites in the first *en* intron, required for early *en* expression²³, were shown to contain binding sites for both Ftz and α Ftz-F1; the two proteins bind to these sites *in vitro* in a cooperative fashion. In the embryo, expression of a minimal reporter gene containing these sites requires the presence of both Ftz and α Ftz-F1, and both DNA binding sites.

The *ftz* enhancer is perhaps the best-characterized target of Ftz activity, and contains α Ftz-F1 binding sites^{7,8}. Recently, Yu *et al.*³² screened in yeast for Ftz-interacting proteins, using portions of the *ftz* enhancer, and isolated α Ftz-F1. The two proteins bind cooperatively to the sites used in their screen. These data support our findings of a direct interaction between Ftz and Ftz-F1, but are surprising in that α Ftz-F1 mutations have no obvious effects on *ftz* expression during the stages we examined. Also, we find that Ftz expressed ectopically is still capable of broadening endogenous *ftz* stripes in an α Ftz-F1 mutant background (data not shown). A possible explanation for these apparent discrepancies is that our α Ftz-F1 mutations affect only a subset of α Ftz-F1 activities. However, this possibility is not consistent with our genetic and molecular analyses; rather, our results suggest two alternatives. The first is that sequences in the *ftz* promoter bound by α Ftz-F1 *in vitro* are not

occupied by α Ftz-F1 *in vivo*. The second is that these sites are occupied by α Ftz-F1 *in vivo*, but that the contribution of these complexes, in the context of the whole *ftz* promoter, is not essential, owing to the redundant action of other cofactors. Regardless of the explanation, the ability of Ftz to autoregulate in the absence of its homeodomain, and in the absence of either Prd or α Ftz-F1, indicates that there are additional Ftz cofactors yet to be identified.

In summary, α Ftz-F1 is a maternally provided cofactor required for Ftz-mediated regulation of the *en* and *wg* genes. Indeed, the two proteins seem to be mutually dependent cofactors for all processes tested except *ftz* autoregulation. The ability of Ftz to act in the absence of its homeodomain suggests that α Ftz-F1 may be important for recruitment of the Ftz polypeptide to specific DNA sequences. In turn, Ftz may influence the transcriptional activity of Ftz-F1. Ftz-F1 is an orphan nuclear receptor, as no ligand has yet been identified²⁴. Nevertheless, it is tempting to speculate that other nuclear receptors, for which ligands have been identified, might also form complexes with homeodomain proteins. This would provide a plausible mechanism by which hormones could, through their nuclear receptors, modulate the activity of homeodomain proteins. Conversely, interactions of this sort might allow homeodomain proteins to modulate the activity or target range of nuclear hormone receptors. □

Methods

Isolation of *Ftz-F1* alleles. We carried out a P-element mutagenesis²⁵, and screened for maternal-effect lethal mutations affecting the anteroposterior polarity of embryos. Of these mutations, two had indistinguishable pair-rule phenotypes, and after further analysis were designated as *Ftz-F1*²⁰⁹ and *Ftz-F1*²⁸². Both alleles are homozygous viable and maternal-effect lethal. The P-element inserts map to the cytological loci 75D and 67D, respectively. Both fail to complement an overlapping deficiency at 75D (*Df(3L)cat*) but complement a deficiency uncovering 67D, demonstrating that the mutations responsible for the phenotype lie in the 75D locus. *Ftz-F1*²⁸² exhibits the same pair-rule phenotype as *Ftz-F1*²⁰⁹ (data not shown). Embryos from *Ftz-F1*²⁰⁹/*Df(3L)cat* trans-heterozygous females and those from *Ftz-F1*²⁰⁹/*Ftz-F1*²⁰⁹ homozygous females (data not shown) also exhibit indistinguishable phenotypes. The P element in *Ftz-F1*²⁰⁹ was remobilized, generating wild-type, maternal-effect lethal, as well as homozygous lethal, excisions. The defects observed in embryos from females homozygous for *Ftz-F1*²⁰⁹ and *Ftz-F1*²⁸² cannot be rescued by wild-type sperm (data not shown). In addition, homozygous mutant embryos with zygotic lethal excisions exhibit no segmental defects. Hence the pair-rule phenotype observed is strictly maternal.

Zygotic rescue of *Ftz-F1*²⁰⁹. *Ftz-F1*²⁰⁹/*Df(3L)cat* females were crossed with males homozygous for *hs* α Ftz-F1, or with wild-type males (control). Embryos 2 h old were heat-shocked for either 0 or 10 min, and allowed to develop for 24 h before preparation of cuticles²⁰. Cuticles were observed under dark-field

illumination. The *Df(3L)cat* stock was obtained from the Bloomington Stock Center.

Northern blot analysis. Total RNA was isolated from *oreR*, *Ftz-F1*²⁰⁹ and *Ftz-F1*²⁸² homozygous females, resolved on a 1.5% formaldehyde-agarose gel, and blotted onto nitrocellulose. Hybridization was performed at 55°C in 50% formamide, 0.25 M NaPO₄, pH 7.2, 0.24 M NaCl, 0.1 mM EDTA, 7% SDS. 5' Antisense riboprobe was made using a 0.9-kb *SacI* fragment specific for the α cDNA, and subcloned under control of the T3 promoter. Transcription reactions were carried out using a Boehringer Mannheim transcription kit.

Histochemical analysis. Antibody stainings were performed using mouse monoclonal anti-En (ref. 26), mouse monoclonal anti-Wg (provided by S. Cohen), and rabbit anti-Ftz (ref. 27), as previously described²⁸. *In situ* hybridization to *ftz* mRNA was as previously described²⁹.

Affinity chromatography and far western analysis. Ftz affinity chromatography using His-tagged Ftz protein and nickel resins was performed as previously described¹⁵. Far western analysis was performed as follows. Full-length and homeodomain-deleted (Δ 273–303) Ftz polypeptides were expressed in bacteria¹⁵. *TcFtz* constructs were also expressed in bacteria. Proteins expressed *in vitro* were made using a Promega TNT *in vitro* transcription/translation kit. The N-terminal Ftz polypeptide was truncated at the unique *Sall* site, and the TcFtz N-terminal polypeptide at the unique *SacI* site. Blotting of SDS-PAGE gels was as described³⁰. To make the α Ftz-F1 probe, a full-length α Ftz-F1 cDNA⁵, under control of the phage T7 promoter, was expressed *in vitro* as described above, except with ³⁵S-methionine added to label the protein. Blots were incubated with 50 μ l labelled α Ftz-F1 in 5 ml cocktail (0.1 M NaCl, 20 mM Tris, pH 7.6, 1 mM EDTA, 1 mM DTT, 10% glycerol, and 1% milk powder) for 2 h at 4°C. The filter was washed in the above buffer for 1.5 h at 4°C, dried and autoradiographed.

Ectopic-expression studies. P-element vectors expressing *ftz* deletion derivatives under *hsp70* promoter control were generated in the vector pNMT4 (ref. 14). Lines expressing Ftz Δ 101–150 have been previously described¹⁵. Embryo collections, heat shocking, *in situ* hybridization and cuticle preparation protocols have all been previously described¹⁴.

Received 24 September; accepted 28 November 1996.

1. Laughon, A. & Scott, M. P. *Nature* **310**, 25–30 (1984).
2. Wakimoto, B. T. & Kaufman, T. C. *Dev. Biol.* **81**, 51–64 (1981).
3. Lavorgna, G., Ueda, H., Clos, J. & Wu, C. *Science* **248**, 848–851 (1991).
4. Lavorgna, G., Karim, F., Thummel, C. & Wu, C. *Proc. Natl Acad. Sci. USA* **90**, 3004–3008 (1993).
5. Ohno, C. K. & Petkovich, M. *Mech. Dev.* **40**, 13–24 (1992).
6. Woodard, C., Baehrecke, E. & Thummel, C. *Cell* **79**, 607–615 (1994).
7. Ueda, H., Sonoda, S., Brown, J. L., Scott, M. P. & Wu, C. *Genes Dev.* **4**, 624–635 (1990).
8. Han, W., Yu, Y. & Pick, L. *Mol. Cell. Biol.* **13**, 5549–5559 (1993).
9. Tsai, C. & Gergen, P. *Development* **121**, 453–462 (1995).
10. Kornberg, T., Sidén, I., O'Farrell, P. & Simon, M. *Cell* **40**, 45–53 (1985).
11. Baker, N. E. *EMBO J.* **6**, 1765–1773 (1987).
12. DiNardo, S. & O'Farrell, P. H. *Genes Dev.* **1**, 1212–1225 (1987).
13. Ingham, P. W., Baker, N. E. & Martinez-Arias, A. *Nature* **331**, 73–75 (1988).
14. Fitzpatrick, V. D., Percival-Smith, A., Ingles, C. J. & Krause, H. M. *Nature* **356**, 610–612 (1992).
15. Copeland, J., Nasiadka, A., Dietrich, B. & Krause, H. M. *Nature* **379**, 162–165 (1996).
16. Jost, W., Yu, Y., Pick, L., Preiss, A. & Maier, D. *Wilhelm Roux Arch. Dev. Biol.* **205**, 160–170 (1995).
17. Brown, S. J., Hilgenfeld, R. B. & Denell, R. E. *Proc. Natl Acad. Sci. USA* **91**, 12922–12926 (1994).
18. Dawes, R., Dawson, I., Falciani, F., Tear, G. & Akam, M. *Development* **120**, 1561–1572 (1994).
19. Stewart, J. J., Brown, S. J., Beeman, R. W. & Dennell, R. E. *Nature* **350**, 72–74 (1991).
20. Ish-Horowicz, D., Pinchin, S. M., Ingham, P. W. & Gyurkovics, H. G. *Cell* **57**, 223–232 (1989).
21. Struhl, G. *Nature* **318**, 677–680 (1985).
22. Carroll, S. B. & Scott, M. P. *Cell* **45**, 113–126 (1986).
23. Kassisi, J. A. *Genes Dev.* **4**, 433–443 (1990).
24. Mangelsdorf, D. J. & Evans, R. M. *Cell* **83**, 841–850 (1996).
25. Erdélyi, M., Michon, A.-M., Guichet, A., Glotzer, J. B. & Ephrussi, A. *Nature* **377**, 524–527 (1995).
26. Patel, N. H. *et al.* *Cell* **58**, 955–966 (1989).
27. Krause, H. M., Klemenz, R. & Gehring, W. J. *Genes Dev.* **2**, 1021–1036 (1988).
28. Ephrussi, A. & Lehmann, R. *Nature* **358**, 387–392 (1992).
29. Tautz, D. & Pfeifle, C. *Chromosoma* **98**, 81–85 (1989).
30. Staudt, L. M. *et al.* *Science* **241**, 577–580 (1988).
31. Florence, B., Guichet, A., Ephrussi, A. & Laughon, A. *Development* **124**, 839–847 (1997).
32. Yu, Y. *et al.* *Nature* (in press).

Acknowledgements. A.G. and J.W.R.C. are first coauthors, and H.M.K. and A.E. are senior coauthors. We thank A. Cyrklaff for cytological localization of P elements; M. Zeidler for the *hs* α Ftz-F1 transgenic flies; H. Stunnenberg and C. Schulz for the α Ftz-F1, hairy, runt, even-skipped and ftz plasmids; S. Cohen and C. Wu for the Wg, En and Ftz-F1 antibodies; M. Petkovich for α Ftz-F1-expressing constructs; H. Ueda, A. Laughon, B. Florence, M. Petkovich and L. Pick for exchange of unpublished data; S. Cohen and F. Stewart for critical reading of the manuscript; and C. Nusslein-Volhard and colleagues for their participation in screening our female sterile *Drosophila* collection. A.G. was supported by fellowships from the Ministère de la Recherche et du Travail and the EMBL; M.E. by an EMBO long-term fellowship; J.W.R.C. was partly funded by a University of Toronto open fellowship, and D.H. by an NSERC summer student fellowship. The balance of funding in the lab of H.M.K. was provided by the Medical Research Council of Canada. Research in the laboratory of A.E. was funded by the European Molecular Biology Laboratory.

Correspondence and requests for materials should be addressed to H.M.K. (e-mail: h.krause@utoronto.ca) or A.E. (e-mail: ephrussi@stork.Embl-Heidelberg.de).

The nuclear hormone receptor Ftz-F1 is a cofactor for the *Drosophila* homeodomain protein Ftz

Yan Yu*, Willis Li†, Kai Su*, Miyuki Yussa*, Wei Han*, Norbert Perrimon† & Leslie Pick*

* Brookdale Center for Molecular Biology, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, New York 10029, USA

† Howard Hughes Medical Institute, Department of Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA

Homeobox genes specify cell fate and positional identity in embryos throughout the animal kingdom¹. Paradoxically, although each has a specific function *in vivo*, the *in vitro* DNA-binding specificities of homeodomain proteins are overlapping and relatively weak. A current model is that homeodomain proteins interact with cofactors that increase specificity *in vivo*^{2,3}. Here we use a native binding site for the homeodomain protein Fushi tarazu (Ftz) to isolate Ftz-F1, a protein of the nuclear hormone-receptor superfamily and a new Ftz cofactor. Ftz and Ftz-F1 are present in a complex in *Drosophila* embryos. Ftz-F1 facilitates the binding of Ftz to DNA, allowing interactions with weak-affinity sites at concentrations of Ftz that alone bind only high-affinity sites. Embryos lacking Ftz-F1 display ftz-like pair-rule cuticular defects. This phenotype is a result of abnormal ftz function because it is expressed but fails to activate downstream target genes. Cooperative interaction between homeodomain proteins and cofactors of different classes may serve as a general mechanism to increase HOX protein specificity and to broaden the range of target sites they regulate.

ftz is a segmentation gene of the pair-rule class located in the Antennapedia (Antp) complex⁴. Although its homeodomain and *in vitro* binding specificity is very similar to other Antp-class proteins^{5,6}, its role in embryos is unique: loss-of-function *ftz* mutations produce deletions of even-numbered parasegments and ubiquitous expression causes an 'anti-*ftz*' phenotype in which odd-numbered parasegments are missing⁷.

To understand the molecular basis of Ftz function *in vivo*, we developed a modification of the yeast two-hybrid system to identify cofactors that modulate its transcriptional activity (Y.Y., J. Hirsch and L.P., manuscript in preparation). This screen used a native Ftz-target element from the upstream regulatory region of the *ftz* gene itself^{8,9}. The *ftz* proximal enhancer is required to establish and maintain the *ftz* seven stripes¹⁰. A core 323-base-pair proximal enhancer (323-bp fPE; Fig. 1a) contains five native binding sites for Ftz protein that mediate autoregulation^{5,11}. The 323-bp fPE was fused upstream of the yeast *HIS3* gene and integrated into the yeast genome. This reporter gene was expressed at low levels in yeast cells, allowing growth in low concentrations of 3-aminotriazole (3-AT; Fig. 1b). Expression of Ftz did not significantly increase reporter gene expression, enabling Ftz-interacting proteins to be isolated by growth selection in high concentrations of 3-AT. The native Ftz-target element facilitated the isolation of cofactors whose interactions with Ftz protein require their binding to DNA. One complementary DNA was isolated that supported only limited growth in 25 mM 3-AT in the absence of Ftz but robust growth when Ftz was present. At 50 mM 3-AT, little growth was detected without Ftz, but cells grew avidly when Ftz was also expressed (Fig. 1b). This cDNA encodes the full-length open reading frame of the α -form of the nuclear hormone receptor Ftz-F1 (refs 12, 13), originally identified as a DNA-binding protein that interacts with the *ftz* zebra¹² and

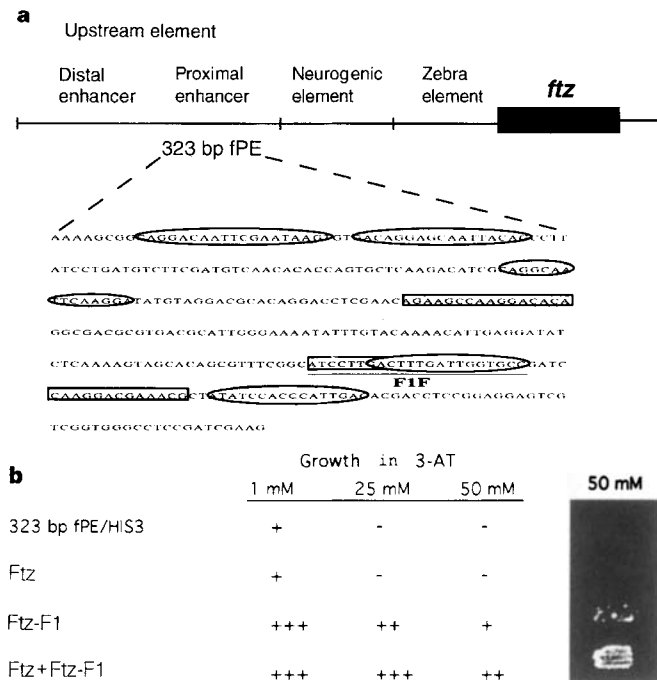


Figure 1 Ftz-F1 was isolated as a Ftz cofactor in a yeast interaction screen. **a**, the *ftz* gene (top); regulatory elements are indicated²⁹. A core region of the proximal enhancer, the 323-bp *ftz* proximal enhancer (323 bp fPE, positions 2,167–2,490) directs expression in seven *ftz*-like stripes *in vivo* through a heterologous promoter¹¹ and contains five binding sites for Ftz protein (ovals); one is a high-affinity, three are of medium affinity and one is a low-affinity site⁵, and there are three binding sites for Ftz-F1 protein (rectangles)¹¹. One of the Ftz-F1 sites abuts a medium-affinity FTZ site (underlined, F1F). **b**, The growth of cells carrying the 323-bp fPE-HIS3 reporter gene in the absence of histidine and in the presence of increasing concentrations of the competitive inhibitor 3-AT. Ftz protein did not significantly stimulate growth; Ftz-F1 supported growth in 25 mM 3-AT and allowed for minimal growth in 50 mM 3-AT. Ftz and Ftz-F1 together strongly activated reporter gene expression, resulting in growth in 50 mM 3-AT.

upstream elements¹¹. As the *ftz-f1* cDNA we isolated was not fused to the Gal4 activation domain, Ftz-F1 is a potent transcription activator in yeast cells, as it is in mammalian and *Drosophila* cells^{14,15}. Our findings suggest that Ftz-F1 not only interacts with DNA but also cooperates with Ftz protein to activate transcription.

Ftz-F1 protein was detected in unfertilized eggs (Fig. 2a), suggesting that it is maternally deposited, and at higher levels during cellularization and germ band extension in all somatic nuclei (Fig. 2b, c). Ftz is expressed in seven stripes that surround the embryo at the cellular blastoderm stage (Fig. 2d). Thus, Ftz and Ftz-F1 proteins are co-expressed in the nuclei of all cells where Ftz is expressed. To test whether Ftz and Ftz-F1 are associated in *Drosophila* embryos, *Drosophila* embryo nuclear extract was immunoprecipitated with anti-Ftz antibodies. Western blot analysis with anti-Ftz-F1 antibody revealed a single 110K band, corresponding to Ftz-F1 (Fig. 2e, lane 1). This band was not detected after immunoprecipitation with preimmune serum (Fig. 2e, lane 2). This suggests that Ftz and Ftz-F1 proteins are complexed *in vivo*.

The 323-bp fPE contains five Ftz binding sites and three Ftz-F1 binding sites. One Ftz-F1 site is adjacent to a medium-affinity Ftz binding site (Fig. 1a). These sites are present in the *Drosophila hydei* *ftz* upstream element¹⁶, suggesting that they have a conserved function. A 27-bp oligonucleotide spanning this region (Ftz-F1/Ftz composite site, abbreviated to F1F; Fig. 1) interacted with Ftz-F1, generating a single band in gel retardation assays (Fig. 3a, bottom arrow in lane 2). Ftz alone (Fig. 3a, lane 3) did not bind F1F DNA under these conditions, although it bound strongly to a

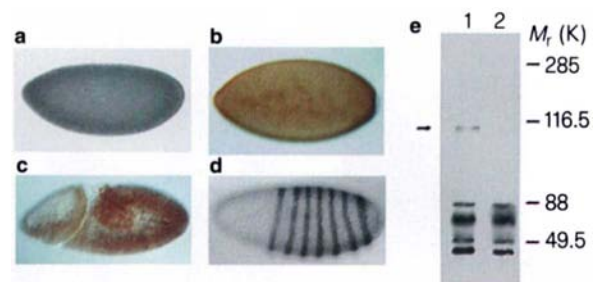


Figure 2 Ftz and Ftz-F1 are coexpressed in developing embryos and from a complex *in vivo*. **a**, Ftz-F1 was detected in unfertilized eggs; **b**, at the cellular blastoderm stage, Ftz-F1 was found in all somatic nuclei but not in the pole cells; **c**, expression remains ubiquitous through germ-band extension. **d**, Expression of Ftz in seven stripes delimits the regions of potential interaction between the two proteins. **e**, Co-immunoprecipitation of Ftz-F1 with Ftz from *Drosophila* nuclear extracts; western blot is shown. Lanes: 1, nuclear extract precipitated with purified anti-Ftz antibody; 2, nuclear extract precipitated with preimmune serum. Arrow indicates the position of Ftz-F1. The four smaller species present in both lanes are nonspecific.

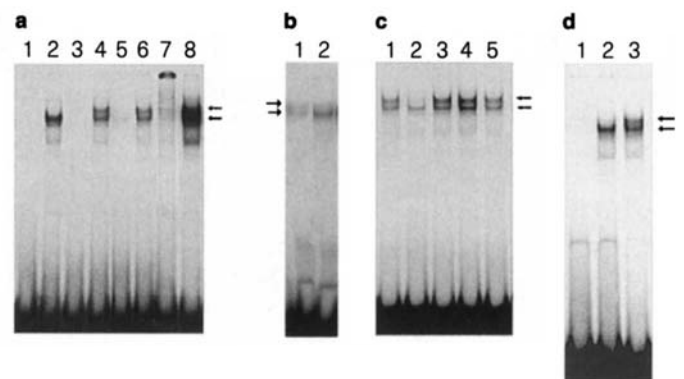


Figure 3 Ftz-F1 facilitates the binding of Ftz to its target DNA. **a**, The binding of Ftz to DNA in the absence and presence of Ftz-F1 was compared in a gel retardation assay. Lanes: 1, control extract; 2, 3 μ g Ftz-F1; 3, 3 μ g Ftz; 4–8, both Ftz and Ftz-F1 proteins. Note the formation of a Ftz/Ftz-F1/DNA ternary complex in lane 4. Binding was inhibited by a 50-fold molar excess of specific competitor (F1F site, lane 5) but not an unrelated oligonucleotide (Slp1 binding site, lane 6). Both the Ftz-F1/DNA complex and the ternary complex were supershifted by anti-Ftz-F1 antibody (lane 7) but not preimmune serum (lane 8). **b**, Anti-FTZ antibody (lane 2) abolished formation of the ternary complex but not the Ftz-F1/DNA binary complex (control, lane 1). **c**, The presence of Ftz in the ternary complex was verified by immunoprecipitation with anti-Ftz antibody (lane 2) compared with preimmune serum (lane 3) or protein A-Sepharose-coated beads alone (lane 4). The ternary complex (lanes 1 and 5, top arrow) was selectively inhibited by anti-Ftz antibody. **d**, Formation of the ternary complex requires both Ftz and Ftz-F1 binding sites. When the core Ftz-F1 binding site was changed from AGGA to ATAG, Ftz-F1 did not bind to DNA and no ternary complex was formed (lane 1). Mutation of the Ftz binding site from ATTG to CGAT had no effect on Ftz-F1 binding to DNA, but the ternary complex was not detected (lane 2). A control reaction with wild-type F1F DNA is shown in lane 3.

high-affinity site at the same concentration (not shown). When both Ftz and Ftz-F1 were included in binding reactions, a more slowly migrating complex formed (Fig. 3a, top arrow, lane 4). The formation of both complexes was specific; a 50-fold molar excess of unlabelled F1F DNA competed for binding (lane 5), but an unrelated oligonucleotide did not (lane 6). This suggested that the slowly migrating complex formed as a result of the specific binding of both Ftz-F1 and Ftz to F1F DNA.

Inclusion of anti-Ftz-F1 antibodies abolished formation of the Ftz-F1/DNA complex and of the more slowly migrating species, whereas preimmune serum had no effect (Fig. 3a, lanes 7 and 8),

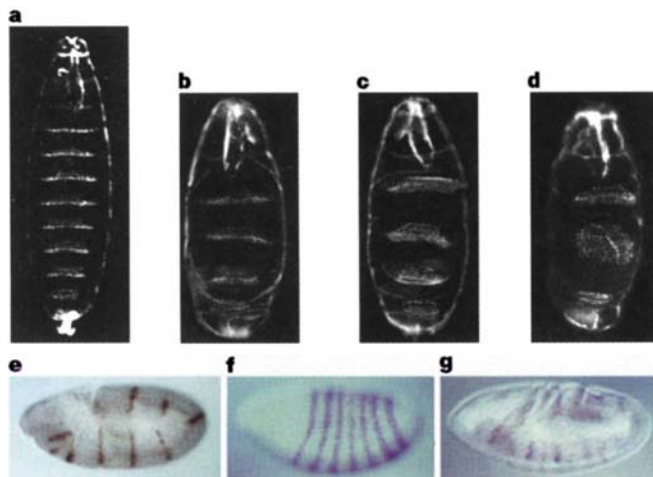


Figure 4 Mutations in *fts-f1* cause *ftz*-like pair-rule defects and alter Ftz function. Cuticle preparations of first instar larva are shown (a–d). a, Wild type, b, *ftz*^{m20} mutant, and c, *fts-f1* mutant, all displaying *ftz*-like pair-rule defects; d, *fts-f1* mutant displaying a more extreme *ftz* phenotype typical of *ftz*^{9H24} mutants. e, Every other En stripe, detected with an anti-En monoclonal antibody, was missing in *fts-f1* mutants. f, The expression pattern of *ftz* RNA, shown at the blastoderm stage and detected with digoxigenin labelled RNA probes, was normal in *fts-f1* mutant embryos. g, The missing En stripes correspond to the Ftz-dependent stripes, such that the remaining stripes (brown) alternate with correctly positioned *ftz* stripes (blue).

indicating that both complexes contain Ftz-F1 protein. Addition of anti-Ftz antibodies abolished formation of the upper complex only (Fig. 3b). To confirm the presence of Ftz in the slowly migrating complex, it was first removed from bacterial preparations by immunoprecipitation. Formation of the slowly migrating complex was selectively reduced (Fig. 3c, lane 2) whereas mock depletion with preimmune serum (lane 3) or protein-A–Sephadex alone (lane 4) had no effect. Together these results show that the slowly migrating complex is a ternary complex containing Ftz-F1, Ftz and DNA. More important, they demonstrate that Ftz-F1 protein dramatically boosts the binding of Ftz to DNA: the ternary Ftz-F1/Ftz/DNA complex formed at concentrations of Ftz that alone did not bind the same oligonucleotide. Ftz alone did not bind to F1F over the 50-fold concentration range tested here, but the ternary complex was detectable at the lowest concentrations of Ftz tested. Therefore the apparent affinity of Ftz for DNA was increased a minimum of 50-fold by Ftz-F1 protein.

To determine whether both proteins interact directly with DNA in the ternary complex, binding was assayed by using oligonucleotides carrying point mutations in either the Ftz-F1 or Ftz binding site within F1F. Mutation of the Ftz-F1 binding site abolished interaction of Ftz-F1 with the oligonucleotide and also abolished formation of the ternary complex (Fig. 3d, lane 1). Mutation of the Ftz binding site had no effect on the formation of a Ftz-F1/DNA binary complex, but the ternary complex was not detected (Fig. 3d, lane 2). Together these results indicate that Ftz and Ftz-F1 bind cooperatively to DNA.

In a large screen for autosomal zygotic-lethal mutations associated with maternal-effect embryonic defects, we identified a P-element-induced mutation, 1(3)03649, which exhibits a pair-rule maternal effect phenotype¹⁷ (Fig. 4). 1(3)03649 mutant animals derived from heterozygous mothers die during larval stages. However, all embryos, irrespective of paternal contribution, derived from homozygous 1(3)03649 germline clones die during embryogenesis and display *ftz*-like pair-rule phenotypes (Fig. 4c, d). 1(3)03649 was mapped by the Berkeley *Drosophila* genome project to 75D4-5. Because the *fts-f1* gene mapped to 75C1-D8 (ref. 13), we

examined whether the P-insertion associated with 1(3)03649 disrupted the *fts-f1* gene. Southern analysis revealed alterations in restriction fragments in the *fts-f1* coding region in 1(3)03649 mutants. We further tested whether the phenotype could be rescued by zygotic expression of a *fts-f1* cDNA under the control of a heat-shock promoter. The maternal-effect phenotype was partially rescued following a 20-min heat shock, and completely rescued following a one-hour heat shock (not shown).

If Ftz-F1 functions *in vivo* as a Ftz cofactor to regulate downstream target genes, *ftz*-like phenotypes would result from loss of Ftz function rather than a change in the expression pattern of the *ftz* gene itself. *ftz* is required for the establishment of alternate *en* stripes. Whereas En protein is detected in fourteen stripes in wild-type embryos, only seven En stripes were detected in *fts-f1* mutant embryos (Fig. 4e), as in *ftz* mutant embryos. In contrast, *ftz* stripes were correctly positioned and expressed at the cellular blastoderm stage in *fts-f1* mutants (Fig. 4f) and stripes remained strong through the stage at which *en* expression is initiated. *ftz* expression was visualized by *in situ* hybridization and embryos were double-stained with anti-En antibodies (Fig. 4g). As expected, the seven remaining En strips were present in alternate parasegment primordia where *ftz* was not expressed. Thus, the Ftz-dependent En stripes were not detectable in *fts-f1* mutant embryos, in support of the idea that Ftz-F1 acts as a Ftz cofactor *in vivo*, because Ftz is present but does not activate *en* expression or direct the development of alternate parasegments in the absence of wild-type Ftz-F1.

Mutations in the *fts-f1* gene mimic loss-of-function *ftz* mutations. Thus, despite the fact that it is expressed ubiquitously in early embryos, loss of maternal Ftz-F1 function results in *ftz*-like pair-rule defects. This suggests that the primary—if not exclusive—function of Ftz-F1 through the cellular blastoderm stage is to serve as a Ftz cofactor. The requirements for Ftz-F1 during later stages of development are unknown. Ftz-F1 binds several sites in *ftz* regulatory regions, including the zebra element¹² and the upstream element¹¹, a region of which was used in our yeast screen (Fig. 1). Mutations in Ftz-F1 binding sites in the zebra element suggested a role in activating *ftz* stripes¹². Similarly, mutations of the three Ftz-F1 binding sites in the 323-bp fPE abolished expression of a *ftz/lacZ* fusion gene in embryos¹⁸, so it was surprising that *ftz* was expressed in *fts-f1* mutants. However, Ftz-F1 may still play a role at late stages in *ftz* autoregulation. First, *ftz* stripes decay slightly earlier in *fts-f1* mutants than in wild-type embryos during germ-band extension, although this effect is subtle. Second, the autoregulatory phase is more clearly observed when enhancer fragments are fused to a basal heterologous promoter^{5,9,10}. Expression of 323-bp proximal enhancer/*lac Z* fusion genes dramatically decreased in *fts-f1* mutants, supporting a role for Ftz-F1 in *ftz* autoregulation (data not shown). Finally, the role of Ftz-F1 in regulating *ftz* expression may be masked because of redundancy. We found that three other *Drosophila* proteins interact with Ftz-F1 binding sites *in vitro*; at least two activate transcription through the 323-bp fPE *in vitro* and are present in early embryos^{11,18}. Therefore, in *fts-f1* mutant embryos, any one of these proteins may be sufficient to activate *ftz* transcription. This is consistent with our previous observations that a high degree of redundancy is built into the *ftz* regulatory network¹⁰.

Genetic analysis shows that Ftz-F1 is required for wild-type function of the *ftz* gene. Ectopic expression defined a 'homeo-domain-independent' function for Ftz^{19–21} which may also be explained by interactions with Ftz-F1 (see accompanying manuscript²²). In contrast, gene rescue experiments using the endogenous *ftz* promoter demonstrated a strict requirement for the homeodomain²³ and we have found that DNA binding is necessary for cooperative Ftz/Ftz-F1 interactions. This indicates that Ftz and Ftz-F1 bind target DNA cooperatively in the wild-type situation, whereas regions of Ftz outside the homeodomain are sufficient to mediate weaker interactions with Ftz-F1 under conditions of ectopic expression.

The cofactor Extradenticle is a homeodomain protein²⁴, which interacts with several HOX proteins to increase DNA-binding affinity and influence binding-site selectivity³. Thus, interactions of HOX proteins with different cofactors will generate different HOX-cofactor complexes that regulate distinct sets of downstream genes. The ability of homeodomain proteins to interact with proteins of the nuclear hormone receptor superfamily increases the interaction combinations that can modulate HOX protein activity in the embryo. Additional control may be provided by binding of Ftz-F1 to an as-yet unidentified ligand. Unlike the *ftz* gene itself, *ftz-f1* is conserved in higher organisms²⁵. The identification of Ftz-F1 as a homeodomain cofactor in *Drosophila* suggests that mammalian Ftz-F1 may have a similar function in regulating target-gene expression in higher organisms. □

Methods

Ftz-F1 was isolated in a yeast double-interaction screen, to be described elsewhere (Y.Y., J. Hirsch and L.P., manuscript in preparation). Immunolocalization of Ftz-F1 was done using standard methods with a rabbit polyclonal antibody provided by C. Wu. For co-immunoprecipitation, affinity-purified rabbit polyclonal anti-Ftz antibody was used. Immunoprecipitation was done with protein A-Sepharose beads (Pharmacia) preincubated with either 10 µg of preimmune serum or purified anti-Ftz antibody. *Drosophila* nuclear extract was prepared from 0–9 h-staged embryos as described¹¹. For western analysis, a 1:1,000 dilution of anti-Ftz-F1 antibody was used. Following incubation with biotin-coupled anti-rabbit IgG, bands were visualized using an ABC Elite kit (Vector Labs). Gel retardation assays were carried out as described¹¹. Binding reactions (25 µl) containing 25 mM HEPES, pH 7.8, 0.5 mM EDTA, 0.5 mM DTT, 10% glycerol, 1 µg poly(dI-dC) and ~10 fmol ³²P-labelled F1F oligonucleotide, and protein, were incubated on ice for 1 h. Reactions were analysed by electrophoresis through 4% non-denaturing polyacrylamide gels using 0.5 × TBE as running buffer. Ftz and Ftz-F1 proteins were synthesized in bacteria with the T7 system, using a pGEMF1 (ref. 26) and pJC20FTZ-F1 (gift from C. Wu) expression plasmids. The *ftz-f1* expression plasmid encodes a protein lacking 200 amino acids at the N terminus and the *ftz* expression plasmid encodes a full-length protein with three amino acids inserted after the initiator methionine. Proteins were prepared as described²⁷ by denaturation with 4M guanidine-HCl, followed by extensive dialysis to facilitate renaturation. Germline clones homozygous for 1(3)03649 were generated using the 'autosomal FLP-DFS' technique²⁸. Briefly, progeny of 1(3)03649 FRT^{3L}/TM3, Sb females were crossed with FLP²²/Y; P(ovo^{D1})^{3L} FRT^{3L}/TM3, Sb. Females were allowed to lay eggs for 1 d in glass vials and their progeny were heat-shocked twice for 2 h at 37 °C in a circulating water bath over a period of two days, when they reached late L2 to L3 larval stages. Subsequently, embryos derived from females of genotype FLP²²; 1(3)03649 FRT^{3L}/P(ovo^{D1}) FRT^{3L} mated with 1(3)03649 FRT^{3L}/TM3, Sb males were analysed. For Southern blot analysis, genomic DNA was digested with *Eco*RI, *Hind*III or *Sac*I. Control genomic DNA was from 1(3)04556/TM3, Sb, a P-element lethal mutation induced on the same parental chromosome as 1(3)03649, but at a different chromosomal location. Zygotic rescue of the maternal effect segmentation phenotype associated with 1(3)03649 was achieved as follows. Embryos derived from females with 1(3)03649 germline clones were crossed to males homozygous for either an *hsp70-ftz-f1* (experimental) or an *hsp70-ftz-f1β/DHR39* (control) transgene (gift from M. Petkovich). The *ftz-f1β/DHR39* gene encodes an orphan receptor that is very similar in structure to Ftz-F1 (refs 14, 15). Embryos of 1.5–2.5 h after egg-laying were heat-shocked by submerging the egg-collection plates, sealed with parafilm, into a 37 °C water bath for 20 or 60 min. The *hsp70-ftz-f1*, but not the *ftz-f1β/DHR39*, transgene was able partially to rescue (*n* = 8/29 embryos) the maternal pair-rule phenotypes associated with 1(3)03649 after 20 min heat shock. Partially rescued embryos showed more than 4 denticle bands. Rescue was complete after 60 min heat shock (of 50 embryos examined, 20 were rescued to various extents, of which 3 appeared to be wild type).

Received 24 September; accepted 18 December 1996.

- McGinnis, W. & Krumlauf, R. *Cell* **68**, 283–302 (1992).
- Hayashi, S. & Scott, M. P. *Cell* **63**, 883–894 (1990).
- Mann, R. S. & Chan, S.-K. *Trends Genet.* **12**, 258–262 (1996).

- Wakimoto, B. T., Turner, F. R. & Kaufman, T. C. *Dev. Biol.* **102**, 147–172 (1984).
- Pick, L., Scheir, A., Affolter, M., Schmidt-Glenewinkel, T. & Gehring, W. J. *Genes Dev.* **4**, 1224–1239 (1990).
- Laughon, A. *Biochemistry* **30**, 11357–11367 (1991).
- Struhl, G. *Nature* **318**, 677–680 (1985).
- Hiroimi, Y. & Gehring, W. J. *Cell* **50**, 963–974 (1987).
- Schier, A. F. & Gehring, W. J. *Nature* **356**, 804–807 (1992).
- Yu, Y. & Pick, L. *Mech. Dev.* **50**, 163–175 (1995).
- Han, W., Yu, Y., Altan, N. & Pick, L. *Mol. Cell. Biol.* **13**, 5549–5559 (1993).
- Ueda, H., Sonoda, S., Brown, J. L., Scott, M. P. & Wu, C. *Genes Dev.* **4**, 624–635 (1990).
- Lavorgna, G., Ueda, H., Clos, J. & Wu, C. *Science* **252**, 848–851 (1991).
- Ohno, C., Ueda, H. & Petkovich, M. *Mol. Cell. Biol.* **14**, 3166–3175 (1994).
- Ayer, S. *et al. Nucl. Acids Res.* **21**, 1619–1627 (1993).
- Maier, D., Preiss, A. & Powell, J. R. *EMBO J.* **9**, 3957–3966 (1990).
- Perrimon, N. A., Lanjuin, C., Arnold, E. & Noll, E. *Genetics* **144**, 1681–1692 (1996).
- Han, W., Yu, Y., Kohanski, R. A. & Pick, L. *J. Biol. Chem.* (submitted).
- Fitzpatrick, V. D. & Ingles, C. J. *Nature* **337**, 666–668 (1989).
- Hyduk, D. & Percival-Smith, A. *Genetics* **142**, 481–492 (1996).
- Copeland, W. R., Nasiadka, A., Dietrich, B. H. & Krause, H. M. *Nature* **379**, 162–165 (1996).
- Guichet, A. *et al. Nature* (this issue).
- Furukubo-Takunaga, K. *et al. Genes Dev.* **6**, 1082–1096 (1992).
- Rauskolb, C., Peifer, M. & Wieschaus, E. *Cell* **74**, 1101–1112 (1993).
- Ikedo, Y. *et al. Mol. Endocrinol.* **7**, 852–860 (1993).
- Krause, H. M., Klemenz, R. & Gehring, W. J. *Genes Dev.* **2**, 1021–1036 (1988).
- Treisman, J., Gonczy, P., Vashishtha, M., Harris, E. & Desplan, C. *Cell* **59**, 553–562 (1989).
- Chou, T.-B. & Perrimon, N. *Genetics* **144**, 1673–1679 (1996).
- Hiroimi, Y., Kuroiwa, A. & Gehring, W. J. *Cell* **43**, 603–613 (1985).

Acknowledgements. We thank R. Kohanski and D. Sassoon for comments on the manuscript; A. Guichet, A. Ephrussi and H. Krause for sharing information before publication; C. Wu for Ftz-F1 antibodies and cDNAs; I. Duncan and S. DiNardo for Ftz and En monoclonal antibodies; S. Elledge for the AACT vector; and M. Petkovich for the heat-shock lines. This is publication 212 from BCMB. N.P. is an Associate Investigator of the Howard Hughes Medical Institute. This work was supported by a grant from the NIH to L.P. L.P. was a recipient of an Irma T. Hirsch career scientist award.

Correspondence and requests for materials should be addressed to L.P. (e-mail: PICK@msvax.mssm.edu).

CORRECTIONS

Structural basis for the binding of a globular antifreeze protein on ice

Zongchao Jia, Carl I. DeLuca, Heman Chao & Peter L. Davies

Nature **384**, 285–288 (1996)

In the legend to Fig. 1b, the orientation of AFP should have been compared to that in Figs 1a, 2b; also, in Fig. 2a legend, the rotation of the planar amide group should have been 180° (not 18° as published). □

Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents

Ravi G. Kurumbail, Anna M. Stevens, James K. Gierse, Joseph J. McDonald, Roderick A. Stegeman, Jina Y. Pak, Daniel Gildehaus, Julie M. Miyashiro, Thomas D. Penning, Karen Seibert, Peter C. Isakson & William C. Stallings

Nature **384**, 644–648 (1996)

We omitted to cite an earlier report on (human) cyclooxygenase-2 structure by C. Luong *et al.*¹. □

- Luong, C., Miller, A., Barnett, I., Chow, J., Ramesha, C. & Browner, M. F. Flexibility of the NSAID binding site in the structure of human cyclooxygenase-2. *Nature Struct. Biol.* **3**, 927–933 (1996).

How to find, identify and quantitate the sugars on proteins

A new sequencer allows the identification, quantification and characterization of sites of glycosylation on proteins, making it possible to analyse the glycosylation at serine and threonine sites in mucin-like domains.

**Andrew A. Gooley and
Keith L. Williams**

The 'one gene—one polypeptide' paradigm is looking outdated as we now understand that in eukaryotes, and even prokaryotes, the output of many genes is post-translationally modified, creating multiple gene products¹. While there are a plethora of co- and post-translational modifications, glycosylation plays a significant role in protein folding, oligomer assembly and protein function. To a significant extent, advances in glycoprotein analysis are limited by inadequacies in the available technologies. Critical issues include understanding what the glycans are, where they are on the protein, and how heterogeneous the structures are at particular sites.

With sugars attached to asparagine residues (the N-linked sugars), potential glycosylation sites are easily located on the polypeptide backbone at the conserved tripeptide motif Asn-Xaa-Ser/Thr/Cys². However, the Asn in the motif is not glycosylated on all secreted/surface proteins and there are specific exclusions. For example, glycosylation does not occur where Xaa=Pro, and inefficient glycosylation can occur when Xaa=Trp, Asp, Glu and Leu³. O-linked sugars, where the sugar is attached to serine or threonine (and occasionally hydroxylysine or hydroxyproline), are more difficult to study since peptide motifs for O-glycosylation are not well defined. In addition, unlike the N-linked sugars where the sugar attached to the amino acid is a conserved N-acetylglucosamine⁴, there are at least seven different sugar linkages and more than one can be found on the same protein⁵ (see Table).

Many think that mass spectrometry holds the key to the identification and analysis of post-translational modifications. Indeed, advances in mass spectrometry are beginning to revolutionize the analysis of protein primary structure, yet sequence analysis of heavily glycosylated regions will always be challenging. Mass spectrometry measures the weight of compounds and many sugar structures are composed of different sugars with the same mass (for example, the N-acetylhexosamines GalNAc and GlcNAc are not distinguished by mass spectrometry). Hence, other techniques must be used in glycoprotein analysis for full structure determination.

We have taken a different approach to address glycosylation site identification.

Figure 1 The GlycoSite cartridge is a screw-cap assembly which contains two teflon pieces in which the glycopeptide is placed, either absorbed on to a membrane of glass fibre (1a, 1b) or covalently coupled to derivatized CPG (controlled pore glass) or PVDF (1c). Edman degradation proceeds normally through glycopeptides until the glycosylated amino acid occurs in the N-terminal position. Although the glycosylated amino acid is cleaved from the peptide, it is poorly extracted by the non-polar solvents (ethyl acetate or chlorobutane) typically used in absorption phase Edman chemistry and most of the glycosylated amino acid remains on the membrane (1a). If polar solvents (such as trifluoroacetic acid, TFA) are used to extract glycosylated amino acids, the whole peptide is washed from the membrane (1b). The GlycoSite is configured so that a single-step delivery of TFA to the cartridge quantitatively extracts the glycosylated and normal protein amino acids that have been released from glycopeptides covalently linked to a membrane support (1c). The extracted glycosylated amino acids can now be identified by HPLC.

Edman degradation involves the stepwise removal of amino acids, whether they are modified or not, starting from the N-terminus of a peptide. In conventional protein sequencing, glycosylated amino acids are not recovered and so there is a blank in the sequencing cycle (Fig. 1a). If glycosylated amino acids are to be extracted following their release from the polypeptide by the Edman chemistry, one needs to use a relatively polar solvent as sugars are polar molecules. With conventional protein sequencing, the protein is only absorbed on to a membrane support, and with a polar solvent the protein is extracted along with the glycosylated amino acid, terminating the sequencing (Fig. 1a & 1b). By chemically coupling the peptide, preferentially at the C-terminus, one can use polar solvents to remove the released glycosylated amino acids and recover them (Fig. 1c).

We first worked with the now defunct Milligen ProSequencer and demonstrated the principle in a paper published in 1991 (ref. 6). In that report a highly glycosylated mucin-like domain (a linker glycopeptide from a *Dictyostelium discoideum* cell-surface glycoprotein) was sequenced⁷. The glycoprotein sequencing technology was subsequently validated on several other glycoproteins^{8–10}.

The next step, with Beckman Instruments of Australia and the United States, was to develop this technology into an instrument, the GlycoSite. GlycoSite separation of the amino acids required that the normal protein amino acids were resolved from the glycosylated amino acids so that both amino acids and

glycosylated amino acids could be identified unambiguously (Fig. 2a). Even with standard sequencing chromatography¹¹, it is possible to predict the nature of the glycoform from its characteristic migration in the liquid chromatography system (Fig. 2b & 2c). For example, it is clear that the glycoforms Ser(diHex), Ser(Hex) and Ser(HexNAc) elute quite differently and have characteristic positions¹² (Fig. 2c). Analysis of 15 glycopeptides/glycoproteins, involving 98 sites of glycosylation, from a diverse range of eukaryotes is reported here (see Table).

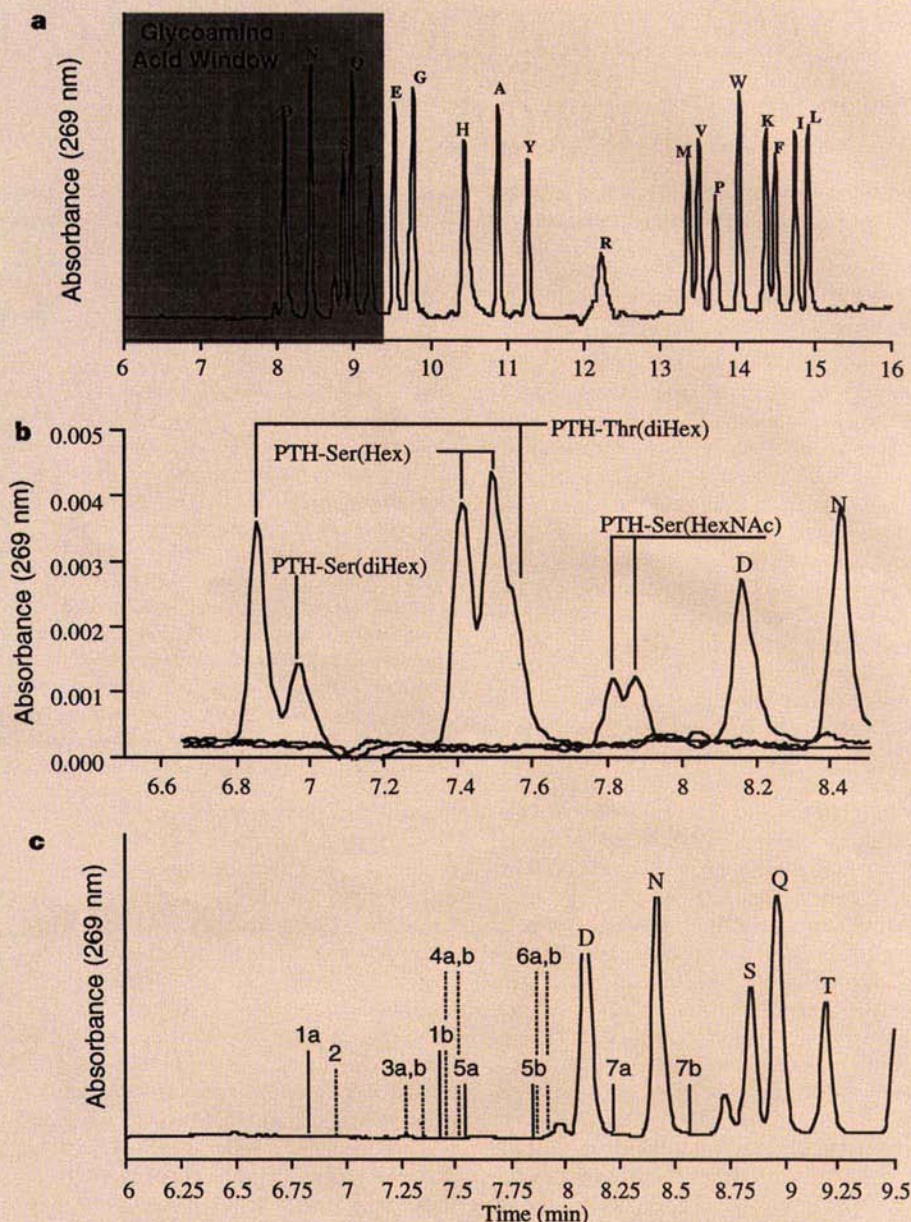
While we have not had problems differentiating between serine- and threonine-linked glycosylation, a generic technology for sugar analysis requires some special considerations as a number of sugar linkages are acid labile. Developments are in progress to recover the acid-labile glycoforms such as fucose, sialic acid and phospho-diester-linked sugars.

One issue to be resolved is that the solid-phase attachment of the glycopeptide is via the carbodiimide activation of carboxyl residues¹³. One must therefore remove sialic acid (which has a free carboxyl group) from oligosaccharide structures, or the sialylated glycosylated amino acid, while cleaved by the Edman chemistry, remains chemically attached to the membrane and is not recovered. Development of an amidation chemistry should overcome this problem¹³.

Now that it is possible positively to identify sites of glycosylation, what is the best strategy to identify glycopeptides? Experienced protein chemists will realize that it is easy to sequence at the N-terminus provided it is not

Figure 2a The chromatography for the PTH-amino acids extracted from the prototype GlycoSite¹¹. An important feature of the GlycoSite chromatogram is that the chromatographic space where the glycoamino acids elute is earlier than that of their parent amino acids Asn, Ser and Thr. All normal amino acids are separated with this chromatography. The chromatogram has been optimized so that even a small percentage of glycosylation (5%) can be detected, as no glycoamino acids have been found to coelute with any of the common protein amino acids. 2b An enlargement of the reversed-phase chromatograms from the GlycoSite analysis of the glycopeptide linker from a recombinant endoglucanase expressed in *Aspergillus*. The chromatogram shows a combined pool of PTH-glycosylated amino acids typical of the linker region, PTH-Thr(diHex), PTH-Ser(diHex) and PTH-Ser(Hex). The standard amino acids Asp and Asn are shown with a PTH-Ser(GlcNAc) from a GST fusion protein expressed in *Dictyostelium*. The Ser modified with an acetylated hexosamine elutes later than the same amino acid modified with a single hexose, which in turn elutes later than a Ser modified with a disaccharide Hex-Hex. The glycoamino acids elute as diastereomers. Our understanding of the diastereomers is that during the Edman degradation the Thr(Sac) and Ser(Sac) are racemized at the β position (where the substituted OH group is located). As a result a mixture of the two Ser and Thr configurations is obtained. These have identical properties in an achiral environment (such as an HPLC column) and one would not expect to separate them. However, when you have a chiral substituent, such as a HexNAc you have two diastereomers, rather than two enantiomers. These now have different properties and resolve on the reversed-phase HPLC column. 2c The elution times for the different sugars we have identified attached to Ser and Thr. Elution profiles are: solid lines, Thr-glycosylated amino acid; dotted lines, Ser-glycosylated amino acid.

1a,b Thr-Hex-Hex 2 Ser-Hex-Hex
5a,b Thr-GalNAc-Gal 3a,b Ser-Hex
7a,b Thr-HexNAc 4a,b Ser-GalNAc-Gal
6a,b Ser-HexNAc



We have not sufficiently resolved the second diastereomer of Ser-Hex-Hex, and we assume that both diastereomers elute as a broad peak at this position. As in the mass spectrometry, the GlycoSite cannot resolve sugar isomers, for example both Ser-GalNAc and Ser-GlcNAc elute at the same position. Confirmation of the HexNAc isomer was achieved by monosaccharide analysis¹⁶.

blocked, but few glycosylated amino acids are found at the N-terminus. Here, having access to a mass spectrometer is enormously beneficial. Steve Carr and colleagues^{14,15} developed a mass spectrometry strategy where glycopeptides and glycoproteins are rapidly identified by their characteristic carbohydrate-specific fragment ions.

Once identified, the glycopeptide can now be analysed by GlycoSite and the sites of glycosylation easily identified and quantified because of the high recovery of the labelled glycosylated amino acid. With this technology we have been able to sequence not only proteins with small amounts of sugar on them but heavily glycosylated regions as well¹².

A major hurdle that remains is a reliable means of sequencing the oligosaccharide. Structures are often inferred simply from their mass and, while this has proved satisfactory for some mammalian N-linked sugars, it does not represent a generic approach for the simple reason that most oligosaccharides are a combination of hexoses and N-acetylhexosamines. An important feature of the GlycoSite technology is that the sugar is labelled with a sensitive ultraviolet absorbing chromophore, the phenylthiohydantoin (PTH)-amino acid, which allows for the accurate quantitation of the recovered glycosylated amino acid (Fig. 3).

We have also shown that the PTH-

glycosylated amino acid is suitable for mass spectrometry, monosaccharide analysis and exoglycosidase digests, so it is applicable to the complete analysis of site-specific glycosylation^{12,16}. Further developments on mass spectrometer-compatible Edman reagents that can be used in the GlycoSite, for example PITC-311 (ref. 17), promise to deliver a highly sensitive approach to the complete characterization of the oligosaccharides attached to glycopeptides.

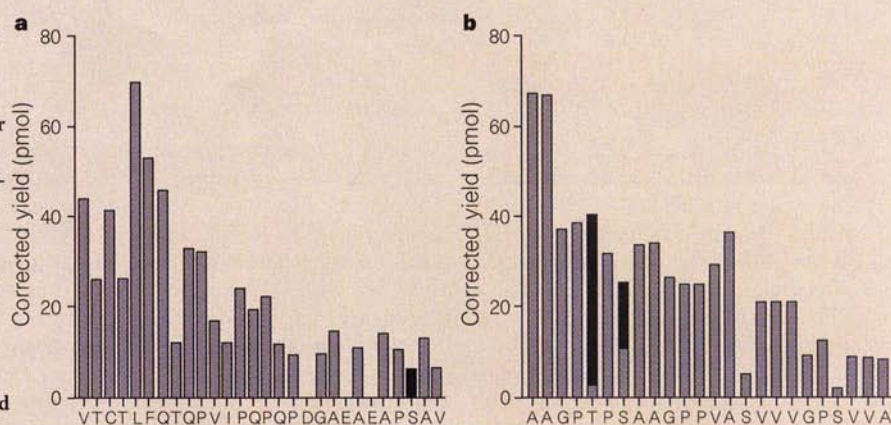
The GlycoSite approach to site identification is sensitive (pmol) and it indicates which sites require further study. The chromatography even gives clues to what kind of sugars are attached to the protein (Fig. 2). It

Table 1 Sites of O-glycosylation identified on the prototype GlycoSite

Sugar linkage	Protein	Number of sites X-Y-Z ⁱⁱⁱ	Glycosylation ^{iv} low-med-high	Ref.
O-Thr (GalNAcGal) ^x	human Glycophorin A ⁱ	9-10-55	OL-OM-9H	8
	bovine κ -casein	6-10-63	3L-2M-1H	9
	rat OD8 α -chain	5-5-23	1L-OM-4H	6
	bovine fetuin	1-1-24	OL-OM-1H	12
O-Ser (GalNAcGal) ^x	human Glycophorin A ⁱ	7-11-55	1L-OM-6H	8
	bovine fetuin	3-5-77	1L-1M-1H	12
O-Thr (GalNAc)	<i>Cryptosporidium</i> sporozoite C1 ^{i,ii}	4-6-22	OL-OM-4H	UO
	Ant antimicrobial F1 peptide	1-1-16	OL-OM-1H	UO
O-Ser (GalNAc)	<i>Cryptosporidium</i> sporozoite C1 ^{i,ii}	2-3-22	OL-OM-2H	UO
	<i>Cryptosporidium</i> sporozoite C6 ^{i,ii}	8-8-16	OL-OM-8H	UO
O-Thr (GlcNAc)	<i>Dictyostelium</i> rPsA ^v	13-14-35	OL-OM-13H	10
	<i>Dictyostelium</i> rPsA ^{vi}	2-4-6	OL-1M-1H	UO
	<i>Dictyostelium</i> GST-fusion protein A ^{vii}	4-4-17	OL-OM-4H	UO
	<i>Dictyostelium</i> GST-fusion protein B ^{vii}	1-1-17	OL-OM-1H	UO
	<i>Dictyostelium</i> ATP protein ^{ii,vi,viii}	1-2-22	OL-OM-1H	UO
O-Ser (GlcNAc)	<i>Dictyostelium</i> GST-fusion protein B ^{vii}	2-3-17	OL-OM-2H	UO
O-Ser (GlcXyl)	Human Coagulation Factor IX	1-1-11	OL-OM-1H	5
O-Ser (FucGlcNAcGal) ^x	Human Coagulation Factor IX ^{ix}	1-2-18	OL-1M-0H	5
O-Thr (HexHex)	<i>Aspergillus</i> endoglucanase	9-10-44	OL-OM-9H	UO
O-Ser (HexHex)	<i>Aspergillus</i> endoglucanase	8-12-44	5L-OM-3H	UO
O-Ser (Hex)	<i>Aspergillus</i> endoglucanase	9-12-44	2L-OM-7H	UO
O-Ser (Hex)	lipase ⁱⁱ of the Loculoascomycetes order	1-6-32	OL-OM-1H	UO

Proteins sequenced from the N-terminus; other proteins were digested and glycopeptides selected for study. ⁱProteins sequenced following separation by SDS-PAGE and electroblotted onto PVDF. In these cases the GlycoSite absorption-phase programme was used. ⁱⁱNumber of sites: number of glycosylated sites (X) followed by the number of Ser or Thr (Y) followed by the total number of amino acids sequenced (Z). ⁱⁱⁱThe glycosylation at each site is described as low (5–25%), medium (25–75%) and high (75–100%) which equates to the quantity of glycamino acid recovered in pmol compared to the expected yield for that site if it were 100% glycosylated. ^{iv}Secreted form of PsA lacking the GPI anchor but containing the Pro/Thr-rich spacer. ^vSecreted truncated form of PsA lacking most of the Pro/Thr-rich linker peptide. ^{vi}Jung, E. *et al. Biochemistry* (in the press). ^{vii}Packer, N. *et al. Electrophoresis* (in the press). ^{viii}The amount of glycosylation was difficult to determine for this acid-labile linkage. ^xSialic acid was removed from the oligosaccharide to prevent attachment of the sialic carboxyl to the arylamine membrane during carbodiimide activation. UO: unpublished observation by MUCAB researchers.

Figure 3 Corrected yields for two bovine fetuin glycopeptides subjected to solid-phase sequencing on the GlycoSite. Desialylated glycopeptide was immobilized on to an arylamine derivatized membrane and sequenced on the GlycoSite¹¹. Each bar represents the corrected yield for a particular cycle from (3a) the first 27 cycles of the glycopeptide Val228-Arg288 and (3b) the first 24 amino acids of the glycopeptide Ala258-Asp292. A single glycosylation site at Ser253 in cycle 25 was recovered with a yield of at least 50%. Initial yield was 40% and repetitive yield was ~94%. Two O-linked glycosylation sites were identified in glycopeptide Ala258-Asp292 at Thr262 (cycle 5, 80% glycosylated) and Ser264 (cycle 7, 50% glycosylated). Initial yield was 50% and repetitive yield was ~94%. The light shaded bars are the yield of PTH-amino acids and the dark solid bars are the PTH-glycosylated amino acids. Asp (D) and Glu (E) are not recovered in the GlycoSite when operating in the solid-phase mode since they are covalently bound by their side-chain carboxyls to the arylamine derivatized membrane.



is likely, as the regulatory authorities appreciate the capacity of this technology, that the quality control requirements for recombinant glycoproteins will become better defined. This raises interesting questions for the biotechnology industry concerning the definition of glycosylation. What percentage glycosylation at a particular site constitutes glycosylation, and how much heterogeneity is acceptable?

Andrew A. Gooley and Keith L. Williams (keith.williams@mq.edu.au) are members of the Macquarie University Centre for Analytical Biotechnology which supports the Australian Proteome Analysis Facility at Macquarie University, Sydney, New South Wales 2109, Australia. The authors

acknowledge support for their glycobiology research from the Australian Research Council, the National Health & Medical Research Council, Novo Nordisk (Denmark), Beckman Instruments (California) and Beckman Australia. Special thanks to Phill Isaacs. For more information fill in reader enquiry number 100.

- Wrotnowski, C. *Genet. Eng. News* (15 May) 7–29 (1996).
- Grinnell, B.W., Walls, J.D. & Gerlitz, B. *J. Biol. Chem.* **266**, 9778–9785 (1991).
- Shakin-Eshleman, S.H., Spitalnik, S.L. & Lakshmi, K. *J. Biol. Chem.* **271**, 6363–6366 (1996).
- Jenkins, N., Parekh, R.B. & James, D.C. *Nature Biotechnol.* **14**, 975–981 (1996).
- Harris, R.J. & Spellman, M.W. *Glycobiology* **3**, 219–224 (1993).
- Gooley, A.A., Classon, B.J., Marschalek, R. & Williams, K.L. *Biochem. Biophys. Res. Comm.* **178**, 1194–1201 (1991).
- Gooley, A.A., Marschalek, R. & Williams, K.L. *Genetics* **130**, 749–756 (1992).
- Pisano, A., Redmond, R.W., Williams, K.L. & Gooley, A.A. *Glycobiology* **3**, 429–435 (1993).
- Pisano, A., Packer, N.H., Redmond, J.W., Williams, K.L. & Gooley, A.A. *Glycobiology* **4**, 837–844 (1994).
- Zachara, N.E. *et al. Eur. J. Biochem.* **238**, 511–518 (1996).
- Pisano, A., Packer, N.H., Redmond, J.W., Williams, K.L. & Gooley, A.A. in *Methods in Protein Structural Analysis* (eds Atassi, M.Z. & Appella, E.) 69–80 (Plenum, New York, 1995).
- Pisano, A. *et al. in Techniques in Glycobiology* (eds Townsend, R. & Hotchkiss, A.) 299–320 (Marcel Dekker, New York, 1996).
- Redmond, J.W. *et al. Glycoconjugate Analysis Patent: PCT/AV 94/00 164*.
- Carr, S.A., Hemling, M.E., Bean, M.F. & Roberts, G.D. *Analyt. Biochem.* **63**, 2802–2824 (1991).
- Carr, S.A., Huddleston, M.J. & Bean, M.F. *Protein. Sci.* **2**, 183–196 (1993).
- Gooley, A.A. *et al. in Techniques in Protein Chemistry VI* (ed. Crabb, J.W.) 83–90 (Academic, New York, 1995).
- Bures, E.J. *et al. in Methods in Protein Structural Analysis* (eds Atassi, M.Z. & Appella, E.) 57–68 (Plenum, New York, 1995).

The main focus for this issue is peptides with some items for proteins thrown in as well. Items include a multimode fraction collector, services for custom peptide synthesis and protein immunology.

Peptide and protein epitope tags

From Boehringer Mannheim

Two new antibodies, anti-VSV-G and anti-myc, for the rapid detection, purification and characterization of epitope-tagged proteins

Epitope tagging eliminates the laborious, time-consuming task of producing an antibody to the specific protein being studied. With anti-VSV-G and anti-c-myc, the tag-specific antibody recognizes its epitope even when the peptide sequence is introduced into unrelated proteins by the epitope tagging technique, according to the company. Suitable for a range of applications, the new epitope tags should be of interest to all researchers involved in protein research, including molecular biologists, biochemists, cell biologists and immunologists. Researchers will be able to gather information about proteins that would otherwise be too difficult to purify or too similar to other proteins to be distinguished *in vivo*.

Reader Enquiry No. 101

Model 206 fraction collector

From Gilson

A high-flow, multimode fraction collection system

Designed to collect eluent either manually or in an automatic mode at flow rates up to 800 ml min^{-1} , the model 206 fraction collector is well-suited for unattended, large-scale preparative applications involving large-volume fractions (of several litres). The fraction collector features six collection modes, including manual, time, peak plus time, time programme and peak plus time programme. With so many modes from which to choose, Gilson says that the instrument should serve most fraction collection requirements. Several features of the collector are designed to help fulfill safety requirements found in large-scale preparative applications. The collector's remote keypad allows for easy placement of the unit under a fume hood or in a cold room. A plastic cover protects the collected fractions from environmental contamination when fractions are collected on a rack. Furthermore, after injection of the sample, the detector's signal can be monitored during a selected time window to check whether the chromatographic separation is progressing as expected. If no compound is eluted from the column within the expected retention time window, the collector is stopped. The collector accepts a variety of test tubes and bottles,



Gilson's multimode fraction collector

including up to 108 10-mm test tubes. Ten different racks can be used on the collector along with one thermostatic rack. As with all Gilson collectors, the model 206 has a stationary rack system designed to eliminate jams, drips and spills; uses $x-y$ stepper-motor technology to provide a method that accurately locates the dispense head over the tubes in the rack and moves the head between, not over, filled tubes to avoid contamination; and has internal drive components made of metal designed to last longer and operate more reliably than plastic in both ambient and cold-room environments.

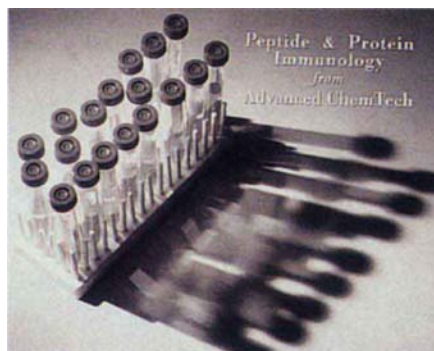
Reader Enquiry No. 102

Peptide antigen synthesis

From Advanced ChemTech

This brochure highlights peptide and protein immunology services

Services include peptide antigen synthesis and preparation, peptide and protein immunochemistry and immunoassay development, as well as antibody production and purification. The brochure details the company's immunology facilities, chemistries and protocols. The company provides custom production of highly sensitive and specific polyclonal antibodies to peptides and proteins, developed to customer specifications and delivered in user-specified time frames. Antisera are delivered according to assay requirements, including diluted and undiluted forms. There are affinity-purified IgGs for researchers engaged in immunoassay, immunoprecipi-



Advanced ChemTech's immunology services

tation, immunoblotting, and other applications where nonspecific reaction minimization is desirable. A variety of high-quality animal hosts are available for antibody projects, including rabbits, guinea pigs, goats and chickens, each with detailed genetic, health and history documentation. A multitude of immunochemistry applications and immunoassay formats can be prepared, including custom antibody-reporter conjugation to enzymes and fluorophores and ^{125}I radiolabelling of peptides, proteins and other compounds. The company will also custom design and produce RIA, ELISA and immunofluorescent assay kits specially prepared and packaged to user specifications.

Reader Enquiry No. 103

Separation and purification

BioSys 2000

From Beckman

Protein purification workstation with separation modelling software

The BioSys 2000 protein purification workstation is a protein method development system that is part of the new ProScale programme. ProScale provides a logical, systematic approach to purification, which is designed to break through major obstacles in the protein drug-development process, helping biopharmaceutical customers to bring products to market faster and more efficiently. The system integrates advanced biochromatography software with high-precision hardware. The new software uses a methodical system of separation strategies, simulating chromatography runs to reduce the trial-and-error nature of protein research. The system works with the HyperD chromatography media and uses an initial set of actual experiments with the user's protein to develop customized simulations. According to Beckman, this simulation saves time and effort by allowing the researcher to explore separation and scale-up strategies without physically running the experiments. The software also builds a database on the protein that can be used at subsequent stages through scale-up. The BioSys 2000 is said to deliver accurate buffer gradients at flow rates of $2 \div 30 \text{ ml min}^{-1}$ to $2 \div 60 \text{ ml min}^{-1}$, with an upper pressure limit of 2,500 p.s.i. A parametric methods designer feature allows users to set up complex multiple-run experiments in minutes by easily manipulating multiple variables. The system has multiple built-in detection capabilities, including ultraviolet/visible, pH and conductivity. BioSys has a multicolumn capacity for maximum flexibility and automatic syringe loading. Moreover, the biocompatible flow path is designed to ensure that there is no loss in biological activity for proteins.

Reader Enquiry No. 104

ÅKTApurifier

From Pharmacia Biotech

A high-performance liquid chromatography system

This is a single-unit, high-performance liquid chromatography system and the second system in the new line of advanced chromatography products. ÅKTApurifier is optimized for purifying peptides, nucleic acids and oligonucleotides using preset running parameters for approximately 80 prepacked columns. The system is controlled and automated by Unicorn software (version 2.20), which has the ability to network other Pharmacia Biotech chromatography systems, regardless of location or scale of operation — from the laboratory bench to full production scale. This common control platform means quick and simple transfer of data from one stage of a project to the next, clear presentation of system status and automatic data evaluation. ÅKTApurifier uses preprogrammed protocols and applications to support all purification techniques. It can also be used for high-resolution separation of protein fragments before sequencing or for mapping purposes. These include high-performance purification of native peptides, separation of protein fragments, routine purification of synthetic peptides and purification of labelled oligonucleotides.

Reader Enquiry No. 105

Spherilose Spherical

From Isco

Spherical, rigid cellulose bioseparations media

This low-pressure media is based on a rigid, spherical, crosslinked cellulose matrix and is designed to offer good stability and flow characteristics. Compared to traditional agarose gels, Spherilose media is said to provide significantly faster separations and reduced cycle volumes, with resultant cost and performance advantages in purifying,



ÅKTApurifier from Pharmacia Biotech

separating and preparing proteins, peptides, polynucleotides and enzymes. The Spherilose sulphate affinity media improves the clean-up of endotoxins and purification of viruses, according to Isco. Spherilose media include 27 different combinations of chemistries and particle sizes for gel filtration, ion exchange, affinity, hydrophobic interaction and adsorption chromatography. All are available in laboratory, pilot and process quantities with Drug Master File (DMF) numbers assigned by the US Food and Drug Administration for most items.

Reader Enquiry No. 106

Microcon-SCX

From Amicon

Microconcentrators for the rapid concentration and purification of peptides and DNA oligomers

New Microcon-SCX adsorptive microconcentrators use a strong cation-exchange membrane to concentrate and remove salts or detergents from peptides, amino acids or DNA oligomers. Adsorption and elution of sample volumes up to 500 µl is said to take only a few minutes using a standard microcentrifuge. Recovery is stated typically to be higher than 90 per cent. Furthermore, up to 400 µg of a peptide may be concentrated, desalted or purified for applications that include HPLC, mass spectroscopy, sequencing and amino acid analysis.

Reader Enquiry No. 107

Sartobind membrane

From Sartorius

Membrane adsorbers for protein concentration, purification and separation using ion-exchange technology

Ion-exchange chromatography is a common technique for the purification and isolation of biomolecules. Currently, macroporous beads are primarily used as a support for functional groups that are active in the exchange of ions. According to Sartorius, the main drawback in operating conventional chromatography columns with bead matrices is that the cycle times depend on the time the target substances take to reach the active functional groups, which are located in the pores of the beads. Other disadvantages of operating chromatography columns are the difficulty of packing and repacking the column and the compressibility of the gel beads. For this reason, conventional chromatography columns can be used only at limited flow rates of 1–10 ml min⁻¹. This also results in long cycle times. Sartorius, therefore, has developed a new separation technology based on microporous membrane ion exchangers. With this technology, Sartorius covalently attaches functional groups to the inner surface of large pore size, modified cellulose membranes. This, the company says, makes these microporous membrane matrices non-compressible and chemically stable.

Reader Enquiry No. 108

Spectral analysis

Dynamo benchtop MALDIMS

Thermo BioAnalysis

Peptide analysis application using MALDI mass spectrometry

A new application note from the company provides practical details on the use of the Dynamo system for rapid analysis of synthetic peptides. Dynamo, a linear, benchtop MALDI MS, includes dynamic extraction for enhanced sensitivity, high resolution

ADVERTISEMENTS

NEWCASTLE UNIVERSITY
MOLECULAR BIOLOGY UNIT



PROTEIN SEQUENCING

Membrane-bound, solid or solution samples

£12 per cycle

5 cycles minimum - no set up charge

For further information contact Dr. Joe Gray
Phone: +44 191 222 8612
Fax: +44 191 222 8129
Email: JOE.GRAY@newcastle.ac.uk

READER ENQUIRY NO. 69

PNA
TIB MOLBIOL
Order per fax...
+49 30 78 79 94-99
or e-mail
DNA@TIB-MOLBIOL.DE

READER ENQUIRY NO. 86

DNA SEQUENCING

£45 per reaction

- PCR products
- M13 inserts
- plasmid inserts
- single strand DNA
- double strand DNA
- no result? no charge!

By research labs for research labs

Molecular Biology Services

King's College School of Medicine and Dentistry
Rayne Institute, 123 Coldharbour Lane, London SE5 9NU
Tel: 0171-346 3126 Fax: 0171-733 3877

READER ENQUIRY NO. 68

and improved mass accuracy. It is said to perform high-resolution analyses at the kind of rapid rates required for continuous monitoring and feedback during the synthesis process. The application note is one of a series describing the use of Dynamo in the analysis of a wide range of compounds of biological, medical and industrial significance. For detailed information on the company's MALDI instruments, contact its new web site (the URL is <http://www.maldi.com>).

Reader Enquiry No. 109

Q-ToF

From Micromass

A new electrospray LC/MS/MS tool for macromolecular characterization, including peptide sequencing

Q-ToF is a new electrospray ionization LC/MS/MS system that should provide a useful alternative to conventional triple-stage quadrupole setups. As such, LC/MS/MS is an enabling technology in the quest to understand the mechanisms of diseases and target effective therapies. Q-ToF is a hybrid MS/MS system, combining the simplicity of a quadrupole [MS 1] with the ultra-high efficiency of an orthogonal time-of-flight (TOF) mass analyser [MS 2]. It exploits TOF MS to achieve continuous production and simultaneous detection of ions across the full mass range. The instrument has a mass range of up to m/z 7,000, in both MS and MS/MS modes, and a resolving power of 5,000 (FWHM), facilitating unambiguous assignment of multiply charged 'peptide sequence' ions. The resolution and stability of the reflectron TOF system delivers good mass measurement accuracy — peptide sequence ions are said to be measured to within ± 0.05 daltons.

Reader Enquiry No. 110

PVA capillary columns for CE

From Hewlett-Packard Europe

A line of capillary electrophoresis columns providing increased longevity, better reproducibility and improved peak shapes

Researchers using capillary electrophoresis (CE) may be interested to learn of a new line of capillary columns, which, at pH 9.0, can typically be used for more than 500 injections. Hewlett-Packard says that reduced electroosmotic flow helps ensure greater reproducibility of retention times, while a reduction in interactions has led to improved peak shape. The PVA capillary columns are intended for the separation of large biomolecules, such as proteins, peptides and glycoproteins, and can also be used to separate small ions and strongly basic molecules, such as amines, for chiral analysis.

Reader Enquiry No. 111

Peptide synthesis

Custom peptides

From Wako BioProducts

Peptides tailor-made to suit individual user requirements

The company now offers peptide synthesis services, as well as other related custom analyses carried out according to desired criteria, which include, but are not limited to, FMOC (tBOC, if desired) or other required chemistries (including solid-phase synthesis). The end-user's specifications for terminal or side-chain manipulations, or additional subsequent analyses (for example, mass spectroscopy), are heeded, with amino acid analysis and a HPLC chromatograph routinely included with each custom peptide order. Peptide agonist/antagonist and drug studies, enzyme inhibitor studies and immunogen synthesis are examples of resultant applications. Furthermore, Wako provides any necessary consultative support. Requests for price quotes should be accompa-

nied by desired amino acid sequence (or number of residues, with any manipulations cited), desired quantity (mg or g amounts) and purity specification (75–80 per cent, >95 per cent or >98 per cent).

Reader Enquiry No. 112

Custom peptide synthesis

From Quantum Biotechnologies

For peptides that are made to order — a complement to any molecular biology toolbox

Four levels of purity are available for custom-made peptides, ranging from crude preparations up to 95 per cent purity. All peptides are synthesized using Fmoc chemistry and researchers can request up to eight standard peptide modifications or, alternatively, a modification of choice. Additional services include special conjugation or modelling studies; the company says it is able to analyse the incidence of any modification on the structure of a peptide.

Reader Enquiry No. 113

Pioneer peptide synthesizer

From PerSeptive Biosystems

A new peptide synthesis system designed to meet a variety of synthesis needs

This system is capable of dual, independent and simultaneous column synthesis, and can be upgraded with a multiple synthesis option to handle the simultaneous synthesis of up to 32 peptides. It is said to combine the possibility for multi-activation chemistries with low reagent consumption, fast cycles, as well as synthesis from small to large scale. The system comes with a full monitoring capability, with protocols for full feedback control to extend cycles automatically when difficult reactions are encountered, fully automated protocols for the synthesis of complex cyclic and branched peptides, and more.

Reader Enquiry No. 114

ADVERTISEMENTS



**PEPTIDE
TECHNOLOGIES**

- Custom Peptide Synthesis
- Purified and Analyzed for Your Research
- Biotinylated, Dye-labelled, Phosphopeptides, etc.
- Peptide Vaccines: KLH Conjugates and MAPS
- CALL 301 869 7306 (USA)
- FAX 301 869 7308
- URL <http://www.luxnet.com/peptide/>

PACEMAKER®

**Assured
Quality**

Custom peptides, conjugates and antibodies

Unrivalled experience and expertise with assured quality

For more information concerning AFFINITI's custom services please contact:

AFFINITI Research Products Ltd., Mamhead Castle, Exeter, EX6 8HD, UK
Tel. +44/0 1626 891010; Fax: +44/0 1626 891090

or visit our new web site: <http://www.affiniti-res.com>

READER ENQUIRY NO. 89

READER ENQUIRY NO. 88